Method for Identifying a Biomolecule

Related Applications

This application is a continuation in part of USSN 09/417,386, filed October 13, 1999, which claims priority to USSN 60/115,109, filed January 8, 1999. Both of these applications are incorporated herein in their entirety.

Field of the Invention

The invention relates to nucleic acids and more particularly to methods of equalizing the representation of nucleic acids in a population of nucleic acid molecules.

Background of the Invention

Approximately 10,000-20,000 genes are thought to be expressed within living cells, depending upon the specific cell type. RNAs corresponding to different genes can be present in different levels in cells. For example, transcripts from as few as 10-15 genes may represent 10-15% of cellular mRNA by mass. In addition to these highly abundant transcripts, another 1000-2000 genes encode moderately abundant transcripts, which can account for up to 50% of cellular mRNA mass. Transcripts from the remaining genes fall into the low abundance class.

Because many genes are identified by isolating complementary DNA (cDNA) corresponding to an RNA sequence, a significant problem can arise because of differences in the levels at which specific RNAs are present in cell types. The most abundant sequences can be repeatedly sampled, while the lowest abundance class may be rarely, if ever, sampled.

Several normalization and subtractive hybridization protocols have been developed to help overcome this problem. These techniques can be technically difficult to perform, and they can fail to detect cDNAs corresponding to rare transcripts.

Summary of the Invention

The invention is based in part on the discovery of novel procedures for equalizing, or normalizing, the representation of nucleic acids in a sample of nucleic acids in which different nucleic acids are initially present in the sample in unequal amounts.

Accordingly, in one aspect the invention provides a method of screening a population of nucleic acid sequences. The method includes providing a population of nucleic acid sequences, partitioning the population into one or more subpopulations of nucleic acids, and identifying a first nucleic acid sequence having an increased level in the subpopulation relative to its level in the starting population of nucleic acids. The first nucleic acid is then compared to a reference nucleic acid sequence or sequences. The absence of the first nucleic acid sequence in the reference nucleic acid or nucleic acid sequences indicates the first nucleic acid is a novel nucleic acid sequence.

The RNA can be derived from a plant, a single-celled animal, a multi-cellular animal, a bacterium, a virus, a fungus, or a yeast. If desired, the RNA can also be partitioned prior to synthesizing cDNA.

Among the advantages of the methods are that they eliminate, or minimize, redundant identification and characterization of identical nucleic acid sequences in a population of nucleic acids..

In some embodiments, the cDNA is synthesized to selectively generate cDNA species that are enriched for those sequences oriented towards the 5'-terminus of the cDNA. In other embodiments, the cDNA is synthesized to enrich for those sequences oriented towards the 3'-terminus of the cDNA.

In some embodiments, the population is normalized by digesting the cDNAs with one or more restriction endonucleases, in different reaction vessels, so as to generate segregated multiple partitions. Preferably, each specific digested cDNA-fragment will occur in only one partition.

In some embodiments, the cDNAs are partitioned by physical methods, which may optionally follow the restriction endonuclease digestion. The physical methods separate the cDNAs a function of their terminal nucleotide sequences, overall length and migratory pattern on a sizing matrix that possesses the ability to separate molecules as a function of their physical and/or biochemical properties.

In other embodiments, the cDNAs are partitioned during subsequent PCR-based amplification of adapter-ligated cDNA fragments that have been digested with one or more restriction endonucleases.

In other embodiments, the cDNAs are partitioned by screening the original mixture of cDNAs so as to remove those sequences that have already been characterized. Screening occurs using partitioned subtraction, whereby the original cDNAs are brought into contact with a prepared, subtraction library of known sequence in such a way that any sequence contained within the original library that is complimentary to any element of the subtraction library is removed or suppressed.

cDNA sequences may also be partitioned by determining the size of each cDNA fragment prior to sequencing; biasing for formation of larger fragment PCR products by lariat formation. In this method, a bias for the larger fragment within the PCR reaction is introduced to allow efficient preferential amplification of longer fragments. Alternatively, partitioning may occur by preferentially amplifying 5' terminal or 3' terminal sequences of mRNA molecules.

If desired, the amplified cDNAs may fractioned by separating the amplified cDNAs on a sizing matrix that separates molecules as a function of their physical and/or biochemical properties and excising individual cDNA fragments from said sizing matrix. The excised cDNA fragments are then inserted into a recombinant vector, or further amplified.

In some embodiments, the restriction endonuclease is a restriction endonuclease that possesses a recognition sequence 4 to 8 basepairs in length and produces either a 5'- or 3-terminal overhang 0 to 6 basepairs in length.

In some embodiments, the identified sequence is subjected to computational analysis. The computational analysis can include querying, or searching, a nucleotide sequence database to identify sequences that match, or the absence of any sequences that match. The database includes a plurality of known nucleotide sequences of nucleic acids that may be present in the sample.

Preferably, the nucleic acid database comprises substantially all the known, expressed nucleic acid sequences derived from a group comprising a plant, a single-celled animal, a multi-cellular animal, a bacterium, a virus, a fungus, or a yeast.

In some embodiments, sizing includes diluting and re-amplification of the cDNAs, fractionating the re-amplified cDNAs by use of one or more sizing matrixes that separate the molecules as a function of their physical and/or biochemical characteristics, physically dividing or cutting the sizing matrixes into a plurality of sections, wherein each section is comprised of one or more cDNAs of similar molecular weight or size. The cDNAs are eluted from each of the sizing matrix section, ligated into a cloning vector and transformed into a host, *e.g.*, a bacterial host. A plurality of the transformed host colonies are selected so as to ensure a statistically-accurate representation of the cDNAs originally contained within the sizing matrix sections. The inserts from this plurality of colonies are recovered and their molecular weight or size of are determined. A plurality of insert DNAs, wherein each successive insert has a molecular weight or size that is within a 0.2 basepair window; and wherein only those DNA species that fall within the 0.2 basepair window is subsequently subjected to nucleotide sequencing.

As utilized herein, the term "normalized" is defined as a mixture of mRNAs (or cDNAs thereof) in which the copy number of highly abundant mRNA species is reduced relative to its copy number in a starting population of nucleic acids, and the copy number of a less abundant mRNA species has been enriched relative to the copy number of the latter mRNA in the starting population.

Among the advantages provided by the present invention are that it multiple partitioning strategies function in a synergistic manner so as to ameliorate unnecessary, redundant sequencing of the same sequence(s), while concomitantly enhancing the sequencing of rarer sequences.

The partition strategies disclosed herein also normalize cDNA abundance by separating the cDNA sequences into multiple partitions possessing minimal sequence overlap. In addition, the various partitioning strategies are performed so as to assure that substantially all cDNAs are sampled. An additional normalization effect may be obtained by separating the resulting DNA fragments based upon their overall size (*i.e.*, size fractionation). Moreover, it is also possible to

normalize the abundance of the cDNAs to an even greater degree by the use of one of several disclosed pre-characterization methods.

In other aspects, the invention pertains to a method of identifiying a nucleic acid sequence by providing a population of nucleic acid molecules including at least one subset of molecules. This subset includes at least one other subset of nucleic acid molecules. The first subset of molecules is then separated and isolated from the rest of the nucleic acid molecules in the population. Next, a library of the isolated first subset of molecules is constructed. One or more members of the library should contain the second subset of molecules, and one or more members of the library should be distinguishable from at least one other member of the library. Subsequently, nucleic acids from one or more members of the library are recovered. Then, the second subset of nucleic acid molecules is separated from at least some of the other members of the library. At least one nucleic acid molecule is separated from the second subset of nucleic acid molecules and is sequenced, thereby identifying a nucleic acid molecule.

In one embodiment, the population of nucleic acids is amplified using a first and a second primer, and, the population of nucleic acid molecules is provided as a plurality of cDNA molecules. These cDNA molecules may be a library including sequences derived from the 5' end of RNA molecules, internal regions of RNA molecules, and/or sequences derived from the 3' end of RNA molecules. Additionally, this library may be amplified using a first and a second primer. In another embodiment, the population of nucleic acid molecules is provided as genomic DNA. In still another embodiment, the population of nucleic acids is a normalized population of DNA.

According to this method of the invention, the first subset of nucleic acids may be separated from the other nucleic acids in the population on the basis of size. Such separation may occur by electrophoresis, e.g., polyacrylamide gel electrophoresis or agarose gel electrophoresis. The members of the first subset of nucleic acids may differ by 20 or fewer nucleotides in length, i.e., 15 or fewer, 12 or fewer, 8 or fewer, or 6 or fewer.

In another embodiment, the population of nucleic acids may also include nucleic acids

having terminal sequences identical to those produced by digestion of a nucleic acid molecule with one or more restriction endonucleases, i.e., a Type II or Type IIS restriction endonuclease. This restriction endonuclease may recognize a nucleotide recognition sequence of varying length. For example, the nucleotide recognition sequence may be 4 or 6 nucleotides in length.

In a further embodiment, the library is prepared by ligating the isolated first subset of nucleic acid molecules to a vector to form a population of vector-insert nucleic acid molecules. These vector-insert nucleic acid molecules are then transformed into a host cell to form a library. Finally, the library may be cultured under conditions to allow for at least some members of the library to be distinguished from other members of the library. In another embodiment, at least one member of the library is spatially distinguishable from other members in the library. Additionally, one or more members of the library may be combined prior to separating the second subset of nucleic acid molecules. In yet another embodiment, the nucleic acid molecule may be compared to one or more known nucleic acid sequences prior to sequencing. The nucleic acids recovered from one or more members of the library may also be pooled prior to sequencing.

In one embodiment, the second subset of nucleic acid molecules may be separated on the basis of size. Such separation may occur by electrophoresis. This electrophoresis may occur in a replaceable matrix formulation that includes a linear polyacrylamide ("LPA") solution, at least one denaturant, a buffer, and 3M to 8M of urea such that the formulation is capable of separating nucleic acids. The LPA concentration may range between 1% to 3% (w/w).

The methods described herein can also be applied to identifying other biomolecules, such as proteins. For example, proteins can be digested with proteases that are specific for different amino-acids and separated on 1-dimensional gels or on 2-dimensional gels using isoelectric

focusing on one dimension and sodium dodecyl sulfate polyacrylamide gel electrophoresis on the second dimension. This step is analogous to the fractionation step of the process as applied to nucleic acids. Following separation, the fragments can be sized by various methods, such as mass spectrometry, to determine suitability for sequencing. The separated and sized peptide fragments can then be sequenced and assembled. By this method, translational profiles of various disease conditions can be elucidated.

All technical and scientific terms used herein have the same meanings commonly understood by one of ordinary skill in the art to which this invention belongs. Although any methods and materials similar or equivalent to those described herein can be used in the practice of the present invention, the preferred methods and materials are now described. The citation or identification of any reference within this application shall not be construed as an admission that such reference is available as prior art to the present invention. All publications mentioned herein are incorporated herein in their entirety by reference.

Brief Description of the Drawings

FIG. 1 is a flow diagram illustrating a method for normalizing the abundance of nucleic acid molecules in a population of nucleic acid molecules.

FIG. 2 is a flow diagram illustrating a method of 5'-enriched cDNA synthesis according to the invention.

FIG. 3A is a schematic diagram showing restriction enzyme digestion and adapter ligation for enrichment of 5' ends of mRNA molecules.

FIG. 3B is a histogram showing the regions of genes covered by clones constructed using 5' end enrichment.

FIG. 3C is a schematic diagram showing restriction enzyme digestion and adapter ligation for enrichment of mRNA molecules containing internal restriction fragments.

FIG. 3D is a histogram showing the regions of genes covered by clones constructed using enrichment for internal restriction fragments.

FIGS. 4A and 4B are schematic illustrations showing the effects of partitioning on the types of nucleic acids recovered in relation to the abundance of the mRNA molecules.

FIG. 5A is a view of a device for separating agarose gels.

FIGS 5B and 5C are representations of front, top, and side view of a device for separating agarose gels.

FIG. 6 is a flow diagram of a system for identifying a nucleic acid.

FIG. 7 is a detailed flow diagram of the process illustrated in FIG. 6.

FIG. 8 is a flow diagram of a system for identifying nucleic acids using pooled-tissues.

FIG. 9 is a detailed flow diagram of the process illustrated in FIG. 8.

FIG. 10 is a flow diagram of a system for identifying a nucleic acid using tissue tagging.

FIGS. 11A, 11B, 11C, and 11D are representations of agarose gels before and after excising regions of the gel containing DNA in the indicated size ranges.

FIG. 12 is a detailed flow diagram of a system for identifying a nucleic acid.

FIG. 13 is a detailed flow diagram for identifying a nucleic acid based on a set of collected traces.

FIG. 14 is a graph showing a plot of actual sequence length of a cloned sequence versus the predicted length of the cloned sequence.

Detailed Description of the Invention

The invention provides methods for identifying nucleic acids in a population of nucleic acid samples. It is based in part on normalizing the representation of sequences that may be initially present in different levels in the population of nucleic acid sequences. The normalization takes place by one or more methods of partitioning the nucleic acid population.

A schematized overview of the invention is shown in FIG. 1. At the input step 100 a starting population of RNA is chosen for analysis. Unless indicated otherwise, reference to a given RNA or population of RNAs is understood to also encompass reference to the corresponding cDNA or cDNAs.

Any population of RNA molecules can be used as long as the population contains, or is suspected of containing, two or more distinct RNA molecules. The population can be isolated

from a starting sample using standard methods for isolating RNA. The RNA population can be isolated from, *e.g.*, an entire organism or multiple organisms, or from a tissue or cell of an organism. The RNA can also be isolated from, *e.g.*, cultured cells, such as eukaryotic or prokaryotic cells grown *in vitro*. If desired, the RNA can be mRNA, (*e.g.*, polyA+ RNA), or stable RNAs (*e.g.*, ribosomal RNA, transfer RNA, or small nuclear RNA). The input RNA or cDNA can be a subpopulation containing the 5' end of RNA molecules (110), a subpopulation having an internal regions of starting RNA molecules (112), or subpopulations containing the 3' end of the cDNA molecules (114).

The selected population or subpopulation is next subjected to a normalization analysis (200). The normalization analysis includes one or more partitioning steps that decrease the relative amount of sequences that are abundant in the starting population of nucleic acids and increase the relative representation of sequences that are rare in the starting population of nucleic acids. A partitioning step can take place before or after mRNA is converted to cDNA. A partitioning step can also take place following amplification of a cDNA. Unless stated otherwise, any partitioning method described herein can be used in conjunction with one or more additional partitioning methods. Examples of suitable partitioning steps are provided below.

In some embodiments, cDNA molecules are subjected to digestion with restriction enzymes, after which adapter oligonucleotides are ligated to the digestion products, and the resulting products amplified. FIG. 1 indicates two types of digestions and adapter ligations which can be performed. The first, designated short chemistry (216) because it tends to result in shorter amplification products, uses two restriction enzymes, followed by ligation of adapter oligonucleotides having termini complementary to the termini of the internal digestion fragments. The second, designated long chemistry (218), similarly uses restriction digestion and adapter ligation but uses longer adapters, which generally result in longer amplification products.

FIG. 1 also illustrates that the modified cDNAs can be subjected to size fractionation (220), which is an example of a partitioning method, and that information from the size fraction analysis can be used in a precharacterization analysis (222). A precharacterization can include, *e.g.*, comparing the size of the insert to sequence databases of fragments sizes produced by the

restriction enzyme. Amplification of short and long chemistry fragments can also be performed in association with partitioning steps, which are explained in detail below.

The amplified products are next sequenced (300). Sequencing can be performed by any method known in the art. The compiled sequence data are then assembled (400), and the sequence generated is compared to known sequences, *e.g.*, sequences in publicly available databases.

Among the advantages provided by the invention are new methods for classifying nucleic acid fragments, *e.g.*, cDNA fragments. by efficiently building data sets of sized and sequenced clones. The methods can also be used for the rapid and efficient generation of databases. In addition, or alternatively, the methods described herein can be applied to 1) genomes that are un-sequenced or genomes with limited sequence information available, 2) tissue specific gene expression profiles, 3) tissue/treatment specific gene expression profiles, or 4) the rapid generation of species and tissue specific databases for an open expression system that relies upon either theoretical restriction digests of known genomes/transcriptomes or custom databases that include the sequences of sized, restriction fragments. Expression systems are described in *e.g.*, U.S. Patent No. 5,871,697 and Shimkets et al., Nature Biotechnology 17:798-803, 1999.

The methods herein described are therefore useful for identifying genes, *e.g.*, expressed genes in an organism of interest, *e.g.*, a human. The sequence information obtained is particularly useful for identifying genes transcribed at low levels, or generating low levels of steady state transcripts. The methods can also be used, *e.g.*, to identify secreted proteins for potential therapeutic use and/or for drug targets; identify variations within the human genome, such as single nucleotide polymorphisms (SNPs); identify differences between normal and diseased tissue; and analyze differential gene expression in different tissues and/or species.

Partitioning prior to cDNA synthesis

One approach to normalize levels of mRNA from a given sample, *e.g.* a given cell or tissue type, is to arbitrarily separate a starting population of RNA molecules into many smaller subpopulations, or collections. In general, a greater number of partitions increases the likelihood that a given partitions will lack a sequence or sequences that is abundant in the starting

population of nucleic acid sequences. This method therefore allows for access to sequences that are expressed in very low copy number.

Alternatively, RNA populations can be isolated from different cell types. This partitioning strategy is based on the premise that different tissues tend to express different subsets of genes. Thus, RNA sequences can be partitioned by sequencing multiple different cDNA libraries extracted from one or more tissues within the body. However, the partitioning will not typically be complete, because many genes are expressed in more than one tissue type.

Synthesis and Amplification of cDNA molecules

Typically, partitioning is performed on cDNA populations that have been modified for subsequent analysis. The modifications may include: (*i*) digesting the cDNA with at least one restriction endonuclease; (*ii*) ligating an adapter oligonucleotide to one or more ends of the termini of the digestion products; and (*iii*) amplifying the ligated products, *e.g.*, in PCR-mediated amplification. These methods are particularly suited to cDNA molecule that have been constructed from the 5', internal, and 3' subpopulation of RNA molecules as described above. These manipulations are collectively known as SEQCALLING™ chemistry. In preferred embodiments, cDNA is generated from populations of RNA molecules that have been divided into subpopulations containing 5' ends of transcripts, populations of molecules containing internal regions of RNA molecules, or subpopulations containing 3' ends of RNA molecules.

A. Construction and amplification of cDNA subpopulation enriched for the 5' ends of mRNA molecules

5'-enriched cDNA synthesis generates cDNA species that are enriched for those sequences oriented towards the 5'-terminus of the cDNA, and in which a specific oligonucleotide sequence is ligated to the 5'-terminus. Approaches for generating cDNAs specifically enriched in transcript 5' ends are often based on the synthesis of a homopolymeric (*e.g.*, dG or dA) tail by the enzyme terminal deoxynucleotidyl transferase (TdT) subsequent to the synthesis of the first cDNA strand. Second strand synthesis is then primed by the use of a complementary homo-oligonucleotide primer sequence. See *e.g.*, Frohman, *et al.*, 1988. *Proc. Natl. Acad. Sci. USA* 85: 8998-9002; Delort, *et al.*, 1989. *Nucl. Acids Res.* 17: 6439-6448; Loh, *et al.*, 1989. *Science* 243:

217-220; Belyavsky, *et al.*, 1989. *Nucl. Acids Res.* 17: 2919-2932; Ohara, *et al.*, 1989. *Proc. Natl. Acad. Sci. USA* 86: 5673-5677.

Alternatively, amplification can exploit the 5'-terminal cap structure present in eukaryotic mRNAs (see *e.g.*, Furuichi & Miura, 1975. *Nature* 253: 374-375; Banerjee, 1980. *Microbiol. Rev.* 44: 175-205; Shatkin, 1985. *Cell* 40: 223-224). However, mRNA preparations generally include a mixture of both capped and non-capped mRNA species. The non-capped mRNAs are thought to be primarily the result of degradation within the cell or during the isolation procedure. An alternative approach to enrich for full-length mRNAs is to purify capped mRNA using affinity reagents. These reagents include naturally occurring proteins that bind the cap structure (see *e.g.*, Edery, *et al.*, 1995. *Mol. Cell. Biol.* 15: 3363-3371); anti-cap antibodies (see *e.g.*, Bochnig, *et al.*, 1987. *Eur J Biochem.* 68: 460-467); and chemical modification of the cap, followed by selection for the modified cap structure (see *e.g.*, Carninci, *et al.*, 1996. *Genomics* 37: 327-336). In addition, 5'-oligo capping can also be used, in which specific oligonucleotide sequences are selectively added to 5'-capped mRNAs prior to first strand cDNA synthesis. Subsequent synthesis of the second strand, is primed by an oligonucleotide that is complementary to the modified cap sequence. See *e.g.*, Maruyama & Sugano, 1994. *Gene* 138: 171-174; Suzyki, *et al.*, 1997. *Gene* 200: 149-156; Fromont-Racine, *et al.*, 1993. *Nucl. Acids Res.* 21: 1683-1684; U.S. Patent No. 5,597,713).

An alternative method for isolating RNA molecules containing a capped 5' end is shown in FIG. 2. FIG. 2 depicts a flow diagram for 5'-enriched cDNA synthesis using a full-length mRNA having a 5'-terminal cap sequence (Gppp) and a poly A+ tail. Also shown in FIG. 2 is truncated mRNA having a 5' terminal phosphate group. Typically, RNA preparations contain a mixture of full-length capped RNAs and truncated mRNAs. The truncated RNAs can arise, *e.g.*, by intracellular degradation of the RNA or by degradation of the RNA during its isolation.

In the first step in FIG. 2, the free 5'-terminal phosphate groups of the truncated or degraded mRNAs are removed by the action of a phosphatase, *e.g.*, the bacterial alkaline phosphatase shown, or calf intestinal alkaline phosphatase. The phosphatase is then inactivated. In the second step, the 5' cap is removed from the full-length mRNA using a pyrophosphatase,

e.g., the tobacco acid pyrophosphatase shown in FIG. 2. The resulting product is the decapped full-length RNA with a free 5'-terminal phosphate group.

In the third step in FIG. 2, the phosphate group serves as a substrate for an RNA ligase-mediated reaction that attaches a specific DNA/RNA hybrid to the 5'-terminus of the full-length mRNAs. An RNA containing the ligated hybrid is used as a substrate for first and second strand cDNA synthesis. Preferably, a combination of oligo(dT)- and random hexamer-mediated first strand priming is performed in the presence of *E. coli* ligase to enhance overall cDNA length. Preferably, an RNase and thermal cycling are used to remove the RNA strand after first strand synthesis. The resulting single strand DNA (ssDNA) functions as a more effective reagent for the priming of second strand synthesis.

Although first strand synthesis occurs for both types of mRNA species (*i.e.,* full-length and truncated/degraded), only those mRNAs with the appropriate sequence ligated to the 5'-terminus (*i.e.,* full-length mRNAs) contain a priming site for subsequent second strand synthesis. Thus, RNAs derived from the full-length mRNAs are selectively amplified.

Preferably, a thermostable enzyme for second strand synthesis in a non-thermal cycled temperature profile is used to ensure more stringent priming of the second strand reaction compared to a non-thermostable enzyme.

A double-stranded cDNA prepared with an adapter containing an oligonucleotide sequence (nR plus "signature sequence") ligated to the 5'-terminus is digested with a restriction endonuclease as shown in FIG. 3A. The oligonucleotide RS (SEQ ID NO:1) (or nR) is used to prime the PCR amplification step subsequent to the ligation of the restriction digestion products. The nJ/nJ PCR product is shown as lined-through to denote that it does not clone efficiently in *E. coli*.

A representation of the distribution of clones derived using 5' enriched synthesis with respect to the region of the gene they include is shown in FIG. 3B. A reference mRNA containing a 5' terminus, an ATG initiation codon, a Stop codon, and a 3' terminus is shown along the X-axis. Also shown is a histogram showing the number of clones (Y-axis) containing sequences derived from the indicated regions of the reference mRNA. The histogram reveals that

the 5' enrichment method method generates distributions enriched in 5' end fragments, and has increased proportions of fragments containing the start codon and the adjacent 90 bp of coding sequences.

B. Construction and amplification of cDNA subpopulations enriched for the interior regions ends of RNA molecules

To generate relatively short cDNA fragments generated from the interior regions of a RNA molecule, i.e., from a region not containing the 5' or 3' terminus, the following procedure is used.

RNA is purified using any standard procedure (see *e.g.,* Berger, 1987. *Methods Enzymol.* 152: 215-219) and cDNA is synthesized according to standard protocols, such as random oligomer or oligo-dT primed synthesis (see, *e.g.*, Gubler & Hoffman, 1983, Gene 25: 263-269, Okayama & Berg, 1982, Mol. Cell Biol. 2: 161-170).

The cDNA is initially digested with a pair of restriction endonucleases. Although any enzyme pair that generates distinct 5'-terminus overhangs is acceptable, a preferred embodiment utilizes enzymes that possess a 4-8 basepair (bp) recognition site yielding a 0-6 bp 5'-terminal overhang, and a more preferred embodiment utilizes enzymes that possess a 6 bp recognition sequence and generates a 4 bp 5'-terminus overhang. One form of manipulation for generating internal fragments is shown in FIG. 3C. The cDNAs are digested with two restriction endonucleases, yielding three types of fragments (two "homo", one "hetero" termini). Following digestion, specific adapters are ligated and the fragments are PCR amplified based upon the specific adapter sequence utilized. As indicated by the crossed lines, the nR--nR and nJ--nJ fragments are unstable in *E. coli*, and are rarely observed following cloning.

Two suitable 24 nucleotide adapter molecules can be generated from RA24 (SEQ ID NO:9); RC24 (SEQ ID NO:10); JA24 (SEQ ID NO:11); or JC24 (SEQ ID NO:12). The adapters are generated by annealing the RA24, RC24, JA24 or JC24 24-mer oligonucleotides (SEQ ID NOs:9-12, respectively) with 12-mer oligonucleotides possessing sequences that are complementary to the last 8 nt of the 3'-terminus of the 24-mer and the 4 bp overhang. The sequences of these primers and other primers described herein are provided in Table 1.

These 4 bp overhang sequences are chosen so as to be complementary to the overhangs that are generated by the restriction endonuclease digestions. In addition, the last 3'-terminal nucleotide of the 24-mer adapter (*i.e.,* A or C) is selected such that a functional restriction endonuclease recognition site is not re-generated when the adapter anneals to the digested cDNA.

Following ligation of the adapters, the restriction endonucleases are heat-inactivated, and the reaction mixture is PCR amplified.

Internal fragments may alternatively be generated using a second type of adapters, which results in longer amplified fragments (also referred to as "Long Internal Chemistry" or "Long Chemistry"). This method is similar to short chemistry, except all adapters possess an additional common sequence on their 5'-termini. This technique suppresses the amplification of small fragments while concomitantly increasing the amplification of longer fragments. The subsequent PCR amplification with the "X" and "J" primers results in production of both a hetero (*i.e.,* "RX--JR") adapter fragment and "homo" adapter fragments (*i.e.,* "RX--XR" and "RJ--JR"), which are unstable in a host and are rarely observed following the cloning process.

The effectiveness of enriching for internal fragments is shown in FIG. 3D. Several thousand sequences generated from internal cDNA framents and compared against a database of approximately 5000 known genes with annotated start and stop sites. Each sequence matching the database was assigned a location on the gene relative to the start (0.0) and stop (1.0) locations relative to the location of the 5'-most matching nucleotide (of the gene). The distribution from a standard run shows that most fragments are located "internally" (*i.e.,* within the coding region). Fragments covering the start codon plus an additional 90 bp (located immediately 3' of the start codon) are significant, because they have a high probability of containing enough sequence to identify secreted proteins. A small but significant fraction of the fragments covers the start codon and the additional 90 bp.

Following digestion, adapters are ligated to these 5'-terminal overhangs. The primers are longer relative to primers used to generate short fragments. Two specific pairs of adapter molecules that can be used in long chemistry synthesis include RXC (SEQ ID NO:2); RXA (SEQ ID NO:3); RJC (SEQ ID NO:4); or RJA (SEQ ID NO:5). The adapters are generated by annealing RXC, RXA, RJC or RJA oligonucleotides (SEQ ID NOs:2-5, respectively) with 12-

mer oligonucleotides possessing sequences that are complementary to the last 8 nt of the 3'-terminus of the 24-mer and the 4 bp overhang. These 4 bp overhang sequences are chosen so as to be complementary to the overhangs that are generated by the restriction endonuclease digestions. In addition, the last 3'-terminal nucleotide of the 24-mer adapter (*i.e.,* A or C) is selected such that a functional restriction endonuclease recognition site is not re-generated when the adapter anneals to the digested cDNA.

Following the ligation of the adapters, the restriction endonucleases are heat inactivated and the reaction mixture is PCR amplified. While the sequences of the two adapters are distinct, they nevertheless possess common 5' sequences that allow the formation of lariat or pan-handle structures that function to suppress PCR-mediated amplification of the shorter fragments.

C. cDNA Synthesis of molecules enriched for 3' ends

3'-enriched cDNA synthesis generates cDNAs that are enriched for the sequences oriented towards the 3'-terminus of the cDNA. This is accomplished by synthesis of the first-strand using a specific oligonucleotide sequence that has been modified to contain an adapter sequence at its 5-terminus (SEQ ID NO:14). Following first-stand cDNA synthesis with the primer, standard cDNA synthesis protocols are utilized as illustrated in FIG. 2.

The 3'-enriched cDNA is digested with one restriction endonuclease. Although any enzyme that generates a distinct 5'-terminus overhang is acceptable, it is generally most preferred to utilize an enzyme that possesses a 6 bp recognition site yielding a 4 bp 5'-terminal overhang. Following digestion, an adapter is then ligated to these 5'-terminal overhangs. These adapters are generated from the JA24 (SEQ ID NO:11) or JC24 (SEQ ID NO:12) 24-mer annealed with 12-mer oligonucleotides possessing sequences that are complementary to the last 8 nt of the 3'-terminus of the 24-mer and the 4 bp overhang. These 4 bp overhang sequences are chosen so as to be complementary to the overhangs that are generated by the restriction endonuclease digestions. In addition, the last 3'-terminal nucleotide of the 24-mer adapter (*i.e.,* A or C) is selected such that a functional restriction endonuclease recognition site is not re-generated when the adapter anneals to the digested cDNA.

Following the ligation of the adapters, the restriction endonucleases are heat inactivated and the reaction mixture is PCR amplified.

Longer fragments enriched for the 3'-ends can be obtained by ligating a longer primer to cDNA molecules that have been digested with a restriction enzyme. Any enzyme that generates a distinct 5'-terminus overhang can be used. It is generally preferred to utilize an enzyme that possesses a 6 bp recognition site yielding a 4 bp 5'-terminal overhang. Following digestion, an adapter is then ligated to the 5'-terminal overhangs. Acceptable adapters are generated from the JA24 (SEQ ID NO:11) or JC24 (SEQ ID NO:12) 24-mer annealed with 12-mer oligonucleotides possessing sequences that are complementary to the last 8 nt of the 3'-terminus of the 24-mer and the 4 bp overhang. These 4 bp overhang sequences are chosen so as to be complementary to the overhangs that are generated by the restriction endonuclease digestion. In addition, the last 3'-terminal nucleotide of the 24-mer adapter (*i.e.,* A or C) is selected such that a functional restriction endonuclease recognition site is not regenerated when the adapter anneals to the digested cDNA.

While the sequences of the two adapters are distinct, they possess common 5' sequences that allow the formation of structures that suppress PCR-mediated amplification of the shorter fragments.

Following the ligation of the adapters, the restriction endonucleases are heat inactivated and the reaction mixture is PCR amplified.

The cDNA fragments prepared as above can be size-fractionated, *e.g.*, electrophoretic fractionation on agarose or polyacrylamide gels, or other types of gels comprised of a similar material. The cDNA fragments may then be physically excised in defined size ranges (*i.e.,* as identified by size makers) and recovered from the excised gel fragments. Additionally, if the quantities of isolated cDNA fragments are low, they can be amplified, *e.g.*, by PCR amplification For example, if the cDNA fragments are generated by Long Internal SEQCALLING™ Chemistry protocol, they are amplified with J23 (SEQ ID NO:6) and X22 (SEQ ID NO:15) primers (either before or after fractionation) prior to cloning, as these cDNAs cannot be efficiently cloned into *E. coli*. Similarly, if the cDNA fragments are generated by Long 5' SEQCALLING™ Chemistry protocol, they can be amplified by J23 (SEQ ID NO:6) and RS

(SEQ ID NO: 1) oligonucleotides (either before or after fractionation) prior to cloning, as these products cannot be efficiently cloned into *E. coli*.

When PCR amplification is used to amplify fragments, conditions are preferentially chosen to minimize non-productive hybridization events. It has been observed that DNA re-hybridization during the PCR amplification process (designated the "Cot effect"; see *e.g.*, Mathieu-Daude, *et al.*, 1996. *Nucl. Acids Res.* 24: 2080-2084) can inhibit amplification. This effect is particularly evident during later PCR amplification cycles, when a substantial concentration of the amplified product has accumulated and the primer concentration has been depleted. As a result, amplification in the later PCR cycles typically follow non-linear dynamics.

By manipulating PCR amplification reaction conditions, it is possible to markedly enhance the "Cot effect", by the insertion of a slow-annealing step in between the denaturation and re-naturation steps in each PCR amplification cycle. The slow-annealing temperature is chosen so as to be above that of the primer-template melting temperature (T_m), but at or above that of the template-template T_m, thus favoring template-template annealing over template-primer annealing. For example, a 85-75°C decrease in temperature at a 10°C/minute gradient can be utilized

Partitioning methods

One or more of the following techniques, or combinations these techniques, can be used to normalize the abundance of RNA (or their cDNA counterpart) species within a given cell or tissue sample.

(i) Partitioning by restriction endonuclease digestion

A cDNA library can be partitioned into many different sets of fragments by digestion with different restriction enzyme pairs. Fragmentation of the same cDNA library with different sets of restriction enzymes, in different reaction vessels, results in segregated multiple partitions, *i.e.*, each specific fragment will occur in only one partition. The digested fragments can be analyzed further, *e.g.*, by direct sequencing, cloning of the digested fragments or sequencing, or one or more of these techniques.

If desired, the cDNA is digested into fragments of a length that is convenient for sequencing. Preferably, multiple different partitions, *e.g.*, 10-100, 20-750, or 50-250 partitions are obtained.

(ii) Partitioning by fragment size or other physical property

Partitioning can also be performed using other separation methods that separate DNA molecules according to their physical characteristics. The methods can include, *e.g.*, separation based on physical and/or biochemical properties (*i.e.*, molecular weight/size, terminal nucleotide sequences, exact migratory pattern, and the like). Separation methods can include, *e.g.*, gel electrophoresis, including agarose or polyacrylamide gel electrophoresis, high pressure liquid chromatography (HPLC), preparative-scale capillary electrophoresis, and similar methodologies.

In one embodiment, unique cDNAs that represent unique (*i.e.*, not previously sequenced) fragments are selected based on their presence in a characteristic restriction enzyme fragment. In this process, a cDNA population is digested with restriction endonucleases, fractionated, and fragments in a desired size range are recovered. The recovered fragments are then ligated to a vector and transformed into an appropriate host, *e.g.*, *E. coli*. Rather that being directly sequenced following the selection process, the DNA fragments are isolated and separated, *e.g.*, sized using one or more sizing matrixes that separate the molecules as a function of their physical or biochemical properties. The embodiment is thus referrred to as "clone sizing". Those recombinant clones that have an insert with characteristics not present in a reference database are determined to contain a unique DNA fragment. Preferably, only unique fragments are subsequently sequenced.

For example, a DNA fragment that is sized in this way possesses two pieces of information that serve as a unique identifier: (*i*) the identity of the restriction endonuclease used to generate the fragment, and (*ii*) the size of the fragment. With these two pieces of information, fragments are picked for subsequent nucleotide sequencing by searching for a specific fragment within a 0.2 basepair window. If a fragment is present in the window, the *E. coli* clone containing the fragment is re-arrayed on a liquid handling robot such as a Tecan Genesis or Packard Multiprobe device, and sequenced. When multiple fragments are present within the 0.2

bp window, only one is selected to be sequenced. Thus, by use of this sizing filter, sequencing of identical fragments is significantly lowered.

By sizing individual fragments and comparing the observed size to previously determined sequences, i.e., using a "sizing filter", only fragments of unique lengths need to be sequenced.

To pre-size large numbers of fragments, the fragments can be initially pooled as a function of their expected size, so as to ensure the any fragment occurs in a minimum of at least three individual pools.

Size fractionation may be accomplished in a number of ways. One commonly utilized method is electrophoretic fractionation on agarose or polyacrylamide gels, or other types of gels comprised of a similar material. The cDNA fragments may then be physically excised in defined size ranges (*i.e.*, as identified by size makers) and recovered from the excised gel fragments. Additionally, if the quantities of isolated cDNA fragments are low, they can be PCR amplified at this stage. For example, if the cDNA fragments are generated by Long Internal SEQCALLING™ Chemistry protocol, described above, they must be amplified with J23 and X22 primers (either before or after fractionation) prior to cloning, as these cDNAs cannot be efficiently cloned into *E. coli*. Similarly, if the cDNA fragments are generated by Long 5' SEQCALLING™ Chemistry protocol, described above, they must be amplified by J23 and RS oligonucleotides (either before or after fractionation) prior to cloning, as these products cannot be efficiently cloned into *E. coli*.

When nucleic acids are separated by size, they are preferably separated by electrophoresis on polyacrylamide or agarose gels. A preferred separation systeim is in high-resolution capillary electrophoresis.

A preferred embodiment for identifying nucleic acids is termed CloneSizing™ separation. CloneSizing™ separation can be performed on any desired population of nucleic acid molecules. Thus, the input nucleic acid molecules for CloneSizing™ separation can come from many potential sources. If a source of input, such as restriction enzyme digest fragments as generated by standard GENECALLING® chemistry (See U.S. Patent No. 5,871,697 and Shimkets et al., Nature Biotechnology 17:798-803 (1999)) or SEQCALLING™ chemistry described above is used, the restriction enzyme information is used, along with the sizing

information, as a unique fragment identifier.

The restriction enzyme fragments can be from either pooled or unpooled samples. In order to have enough material for direct cloning of the DNA, the GENECALLING® fragments are preferably re-amplified prior to fractionation. The re-amplification is preferably performed with the same primer as is used in the original GENECALLING® chemistry. The number of PCR cycles in the reamplification is preferably kept to a minimum in order to keep primer-dimer formation as low as possible.

Restriction enzyme fragments can also be amplied from pooled samples, from multiple tissues, or the same tissue following exposure to different conditions. In some embodiments, pooled samples are re-amplified with specific, signature, primers. A signature primer is a primer that serves to uniquely identify the source of the nucleic acid fragment after the fragment is sequenced.

To mix tissues, the same subsequence from each sample is individually amplified by PCR with a signature primer. After PCR amplification, the DNA can be concentrated (for example, by ethanol precipitation) precipitated and quantified, *e.g.*, by fluorometry. An equal mixture of two or more tagged samples is preferably prepared prior to fractionation. In one embodiment, for a three-sample pool, 0.75 μg of DNA from each tissue/subsequence is mixed with two other similar samples and the mixed population of DNA fragments is subjected to electrophoresis on Metaphor agarose gels.

To increase the throughput of the process, while maintaining the diversity of the tissues, fragments from different tissues are preferably amplified with a mix of signature primers differing by six bases from each other. If desired, multiple tissues, *e.g.*, 2, 3, or 4 tissues, are pooled together. In order to have an accurate representation of the fragments, the maximum number of mixed tissues that is tolerable depends on the concentration of distinct DNA fragments in the MetaPhor® agarose plugs after fractionation. One can mix more tissues that have relatively few distinct DNA fragments, and fewer of the more complex tissues.

Restriction fragments are then separated in an electrophoretic gel system, *e.g.*, a polyacrylamide or agarose gel system. MetaPhor agarose gel based electrophoresis is the preferred method of separation due to the combination of high-resolution in the 200-800 base-pair range, ease of physical fractionation, and ability to perform direct ligation on DNA

fragments eluted from the gel.

Numerous fractionation protocols can be used. In one preferred embodiment, the DNA bands in the re-amplified GENECALLING® chemistry are separated in 48 fractions containing fragments from 500 to 50 base pairs. To achieve this fractionation, two MetaPhor® agarose gels are cast: a 3% and a 4%, on which the fragments from 500 to 220 base pairs and 220 to 50 base pairs, respectively, will be separated. Then, each gel is physically cut in 24 fractions, comprising fragments within approximately 6 to 12 base pair windows. The fractionation is performed with the devise detailed in FIGS. 5A, 5B, and 5C.

After physical fractionation, the gel plugs are arrayed in a 96 well plate. The cDNA contained within the plugs is eluted by centrifuge force.

Next, the eluted fragments are recovered, and introduced into a cellular host, *e.g.*, a bacterial host such as *E.coli*. In one embodiment, a fractionation method, such as electrophoresis in MetaPhor gel, which allows direct ligation of the DNA fragments into a vector, is performed.

For example, the fraction-separated fragments can be collected, and any method known in the art can be used to ligate the eluted DNA fragments into a suitable vector. A preferred method for the direct ligation of the fragments into the vector is TOPO TA cloning. The TOPO TA cloning® vector (Invitrogen) allows very rapid and efficient ligation into a vector carrying the Lac Z gene for rapid detection of fragment insertion. Competent bacteria cells, such as One Shot® TOP10 chemically competent cells (Invitrogen), are then transformed with the ligation mixture and plated onto selective media, *e.g.*, LB Amp Kan XGal.

After incubation to allow for colony growth, the individual colonies (clones) are identified and transferred into a suitable array, *e.g.*, 384-well plates (Genetix Larger volume plate) filled with 50 μL liquid LB Amp Kan 10% glycerol media. The picking is preferably based upon an a system that allows for rapid identification of colonies which have plasmids containing inserts. The screen can be, *e.g.*, the blue/white selection of the Lac Z gene. A standard automated colony-picking robot, such as the Q-Pix from Genetix (UK) is programmed to pick a minimum of zero and a maximum of ninety-six white colonies per petri dish, and deliver the picked colony to one of the 4 sub-plates on the 384-well plate.

The plating of the transformed cells and subsequent colony picking is preferably performed in a systematic method that allows for a colony to be matched to a gel fraction or

fractions from which the insert in the colony is derived. For example, when the plating is performed, each unique fraction (corresponding to a distinct subsequence pair and size window on, *e.g.*, a MetaPhor gel) can be plated onto a separate petri dish. The picking is performed such that each unique fraction is picked into a unique 96-well subplate of the destination 384-well plates.

At this point, the clones can be sized individually, if desired. While individual sizing may be simple and accurate, performing single-clone sizing on a large scale is cost prohibitive.

An alternative method is to size multiple clones simultaneously, *i.e.*, to perform multiplex sizing. Multiplex sizing allows for quick and cost efficient sizing of multiple clones. For example, after the clones are isolated into wells of a 384-well media plate, the individual clones are pooled together into destination 384-well plates. The pooling automation is driven by a pooling map that is generated by an algorithm designed to allow the sized, multiplexed clones to be mapped back to their original location.

In the ideal case of perfect fractionation, the pooling of the clones from different fractions never results in the possibility of confusion over which clones came from which fraction. Real fractionation, however, is non-ideal, and bands will be present in a fraction where they are not expected. This non-ideal fractionation would result in confusion if the pooling where performed in a naïve fashion where ideal fractionation is assumed.

Any confusion created by non-ideal fractionation can be minimized by the use of pooling maps. For example, 48 fractions can be simultaneously collected from 8 subsequences and treated as a pooling unit. These 8 subsequences and 48 fractions result in 384 total fractions, or 96, 384-well plates. The pooling map is based upon parsing the 384-well plates into 96-well subplates. In this way, pooling preserves the clone location determined by the colony picking and takes advantage of the efficiency of 96-tip automation. The pooling map attempts to minimize the placement of fractions from a single subsequence close to one another and thereby eliminate the possibility of identical fragments being placed into the same well. The chance that aberrantly fractionated fragments from different subsequences will size identically is lower than the chance that aberrantly fractionated fragments from the same subsequence will size identically.

A pooling map can include:

SS1 – Fraction 1

SS2 – Fraction 2

SS3 – Fraction 3

SS4 – Fraction 4

SS5 – Fraction 5

SS6 – Fraction 6

SS7 – Fraction 7

SS8 – Fraction 8

SS1 – Fraction 9

SS2 – Fraction 10

SS3 – Fraction 11

.

.

.

SS8 – Fraction 48

To further enhance the opportunity for sizing each fragment and mapping its size back to the clone, each clone can be pooled into 4 wells (each 96-well subplate is pooled into 4 different 96-well destination subplates). In this way, each clone is run in 4 gel lanes. The advantage of redundant sizing is that if one of the lanes fails due to an electrophoresis failure or PCR failure, there are still other opportunities to size the clone. Additionally, the combination of the pooling map strategy with the multiple lanes of sizing electrophoresis further serves as an identifier of the clone identity. After sizing, the results from all 4 lanes of electrophoresis are compared against one another to find the fragment that is present in all the lanes and within the correct fraction window. If bands are not present in all 4 lanes, or if the bands are not within the correct fraction window, a series of deconvolution rules have been defined to determine the probability that the clone is correct.

Pooling can be accomplished on a number of liquid-handling robots, such as the Matrix PlateMate or the Beckman Multimek-96. Samples are preferably processed in blocks of 96 to allow for rapid processing of the samples, and flexibility of programming to accommodate the pooling map.

In one embodiment, the robot is the Matrix PlateMate 96-tip liquid handling robot (Matrix Tech Corp. Nashua, NH). The robot should be programmed to allow for the pooling of the clones with a pooling map, where the pooling map directs the pooling of the clones in such a manner as to allow correlation between the determined sizes and the location of the clones on the 384-well clone plates.

A preferred method of programming the robot is to drive the PlateMate software from an external, flexible software package that is capable of reading a file such as Visual Basic (Microsoft) ("VB"). In this method, the pooling map is read by the VB program, and the VB program loads small PlateMate application programs in a sequence prescribed by the pooling map into the PlateMate. The PlateMate.ini file can be altered by setting the re-initialization parameter from 1 to 0 in order to allow the VB program to control the robot.

The PlateMate application programs are written as flexible units that can be combined in unlimited combinations. For example, programs can be written for;

-Aspiration from each of the 4 subplates on the source plates

-Dispensing into each of the 4 subplates on the destination plates

-Getting source plates

-Putting source plates

-Getting aspiration plates

-Putting aspiration plates

-Tip wash

The complexity of the multiplex pooling arises from the mapping of the DNA fragments back to the clone in which they are found. For the purposes of the discussion, the specific well on the clone plate will be used to mean the clone.

A pooling map can be be used to correlate isolated clones and the sample in which they are inserted. The pooling map thus facilitate sample analysis and processing, and allows a maximum number of clones to be pooled together with a minimum of ambiguity.

While no particular method is requried for generating a pooling maps, a preferred method is to generate a pooling map for the 48 fractions from 8 subsequences is shown below. A preferred pooling map for the 24 fractions from 16 subsequences follows the 48 fractions from 8 subsequences map. The pooling maps are interpreted in the following way: the initial 32 links of

the map define the 32 subplates that will be pooled into. The 6-digit numbers are bar-codes for the destination pooling plates. The subsequent 384 lines of the map detail where the clones from each source subplate will be delivered. As an explicit example, line 32 of the map is: plate 568401 A1 1 9 17 32. This is interpreted as the clones from the A1 subplate will be placed into the following pool subplates: 801231 A1; 801233 A1; 801235 A1; and 801238 B2.

```
pool  801231 A1 1
pool  801231 A2 2
pool  801231 B1 3
pool  801231 B2 4
pool  801232 A1 5
pool  801232 A2 6
pool  801232 B1 7
pool  801232 B2 8
pool  801233 A1 9
pool  801233 A2 10
pool  801233 B1 11
pool  801233 B2 12
pool  801234 A1 13
pool  801234 A2 14
pool  801234 B1 15
pool  801234 B2 16
pool  801235 A1 17
pool  801235 A2 18
pool  801235 B1 19
pool  801235 B2 20
pool  801236 A1 21
pool  801236 A2 22
pool  801236 B1 23
pool  801236 B2 24
pool  801237 A1 25
pool  801237 A2 26
pool  801237 B1 27
pool  801237 B2 28
pool  801238 A1 29
pool  801238 A2 30
pool  801238 B1 31
pool  801238 B2 32
plate 568401 A1 1 9  17 32
plate 568401 A2 2 10 18 29
plate 568401 B1 3 11 19 26
plate 568401 B2 4 12 20 31
```

plate 568402 A1 5 13 21 28
plate 568402 A2 6 14 22 25
plate 568402 B1 7 15 23 30
plate 568402 B2 8 16 24 27
plate 568403 A1 2 9 17 31
plate 568403 A2 3 10 18 28
plate 568403 B1 4 11 19 25
plate 568403 B2 5 12 20 30
plate 568404 A1 6 13 21 27
plate 568404 A2 7 14 22 32
plate 568404 B1 8 15 23 29
plate 568404 B2 1 16 24 26
plate 568405 A1 3 9 17 30
plate 568405 A2 4 10 18 27
plate 568405 B1 5 11 19 32
plate 568405 B2 6 12 20 29
plate 568406 A1 7 13 21 26
plate 568406 A2 8 14 22 31
plate 568406 B1 1 15 23 28
plate 568406 B2 2 16 24 25
plate 568407 A1 4 9 17 29
plate 568407 A2 5 10 18 26
plate 568407 B1 6 11 19 31
plate 568407 B2 7 12 20 28
plate 568408 A1 8 13 21 25
plate 568408 A2 1 14 22 30
plate 568408 B1 2 15 23 27
plate 568408 B2 3 16 24 32
plate 568409 A1 5 9 17 28
plate 568409 A2 6 10 18 25
plate 568409 B1 7 11 19 30
plate 568409 B2 8 12 20 27
plate 568410 A1 1 13 21 32
plate 568410 A2 2 14 22 29
plate 568410 B1 3 15 23 26
plate 568410 B2 4 16 24 31
plate 568411 A1 6 9 17 27
plate 568411 A2 7 10 18 32
plate 568411 B1 8 11 19 29
plate 568411 B2 1 12 20 26
plate 568412 A1 2 13 21 31
plate 568412 A2 3 14 22 28
plate 568412 B1 4 15 23 25
plate 568412 B2 5 16 24 30
plate 568413 A1 8 10 18 31

plate 568413 A2 1 11 19 28
plate 568413 B1 2 12 20 25
plate 568413 B2 3 13 21 30
plate 568414 A1 4 14 22 27
plate 568414 A2 5 15 23 32
plate 568414 B1 6 16 24 29
plate 568414 B2 7 9 17 26
plate 568415 A1 1 10 18 30
plate 568415 A2 2 11 19 27
plate 568415 B1 3 12 20 32
plate 568415 B2 4 13 21 29
plate 568416 A1 5 14 22 26
plate 568416 A2 6 15 23 31
plate 568416 B1 7 16 24 28
plate 568416 B2 8 9 17 25
plate 568417 A1 2 11 18 28
plate 568417 A2 3 12 19 25
plate 568417 B1 4 13 20 30
plate 568417 B2 5 14 21 27
plate 568418 A1 6 15 22 32
plate 568418 A2 7 16 23 29
plate 568418 B1 8 9 24 26
plate 568418 B2 1 10 17 31
plate 568419 A1 3 11 18 27
plate 568419 A2 4 12 19 32
plate 568419 B1 5 13 20 29
plate 568419 B2 6 14 21 26
plate 568420 A1 7 15 22 31
plate 568420 A2 8 16 23 28
plate 568420 B1 1 9 24 25
plate 568420 B2 2 10 17 30
plate 568421 A1 4 11 18 26
plate 568421 A2 5 12 19 31
plate 568421 B1 6 13 20 28
plate 568421 B2 7 14 21 25
plate 568422 A1 8 15 22 30
plate 568422 A2 1 16 23 27
plate 568422 B1 2 9 24 32
plate 568422 B2 3 10 17 29
plate 568423 A1 5 11 18 25
plate 568423 A2 6 12 19 30
plate 568423 B1 7 13 20 27
plate 568423 B2 8 14 21 32
plate 568424 A1 1 15 22 29
plate 568424 A2 2 16 23 26

plate 568424 B1 3 9 24 31
plate 568424 B2 4 10 17 28
plate 568425 A1 7 12 19 29
plate 568425 A2 8 13 20 26
plate 568425 B1 1 14 21 31
plate 568425 B2 2 15 22 28
plate 568426 A1 3 16 23 25
plate 568426 A2 4 9 24 30
plate 568426 B1 5 10 17 27
plate 568426 B2 6 11 18 32
plate 568427 A1 8 12 19 28
plate 568427 A2 1 13 20 25
plate 568427 B1 2 14 21 30
plate 568427 B2 3 15 22 27
plate 568428 A1 4 16 23 32
plate 568428 A2 5 9 24 29
plate 568428 B1 6 10 17 26
plate 568428 B2 7 11 18 31
plate 568429 A1 1 12 19 27
plate 568429 A2 2 13 20 32
plate 568429 B1 3 14 21 29
plate 568429 B2 4 15 22 26
plate 568430 A1 5 16 23 31
plate 568430 A2 6 9 24 28
plate 568430 B1 7 10 17 25
plate 568430 B2 8 11 18 30
plate 568431 A1 2 12 19 26
plate 568431 A2 3 13 20 31
plate 568431 B1 4 14 21 28
plate 568431 B2 5 15 22 25
plate 568432 A1 6 16 23 30
plate 568432 A2 7 9 24 27
plate 568432 B1 8 10 17 32
plate 568432 B2 1 11 18 29
plate 568433 A1 3 13 19 32
plate 568433 A2 4 14 20 29
plate 568433 B1 5 15 21 26
plate 568433 B2 6 16 22 31
plate 568434 A1 7 9 23 28
plate 568434 A2 8 10 24 25
plate 568434 B1 1 11 17 30
plate 568434 B2 2 12 18 27
plate 568435 A1 4 13 19 31
plate 568435 A2 5 14 20 28
plate 568435 B1 6 15 21 25

```
plate 568435 B2 7 16 22 30
plate 568436 A1 8 9  23 27
plate 568436 A2 1 10 24 32
plate 568436 B1 2 11 17 29
plate 568436 B2 3 12 18 26
plate 568437 A1 6 14 20 27
plate 568437 A2 7 15 21 32
plate 568437 B1 8 16 22 29
plate 568437 B2 1 9  23 26
plate 568438 A1 2 10 24 31
plate 568438 A2 3 11 17 28
plate 568438 B1 4 12 18 25
plate 568438 B2 5 13 19 30
plate 568439 A1 7 14 20 26
plate 568439 A2 8 15 21 31
plate 568439 B1 1 16 22 28
plate 568439 B2 2 9  23 25
plate 568440 A1 3 10 24 30
plate 568440 A2 4 11 17 27
plate 568440 B1 5 12 18 32
plate 568440 B2 6 13 19 29
plate 568441 A1 8 14 20 25
plate 568441 A2 1 15 21 30
plate 568441 B1 2 16 22 27
plate 568441 B2 3 9  23 32
plate 568442 A1 4 10 24 29
plate 568442 A2 5 11 17 26
plate 568442 B1 6 12 18 31
plate 568442 B2 7 13 19 28
plate 568443 A1 1 14 20 32
plate 568443 A2 2 15 21 29
plate 568443 B1 3 16 22 26
plate 568443 B2 4 9  23 31
plate 568444 A1 5 10 24 28
plate 568444 A2 6 11 17 25
plate 568444 B1 7 12 18 30
plate 568444 B2 8 13 19 27
plate 568445 A1 2 14 20 31
plate 568445 A2 3 15 21 28
plate 568445 B1 4 16 22 25
plate 568445 B2 5 9  23 30
plate 568446 A1 6 10 24 27
plate 568446 A2 7 11 17 32
plate 568446 B1 8 12 18 29
plate 568446 B2 1 13 19 26
```

```
plate 568447 A1 3 14 20 30
plate 568447 A2 4 15 21 27
plate 568447 B1 5 16 22 32
plate 568447 B2 6 9  23 29
plate 568448 A1 7 10 24 26
plate 568448 A2 8 11 17 31
plate 568448 B1 1 12 18 28
plate 568448 B2 2 13 19 25
plate 568449 A1 5 16 21 25
plate 568449 A2 6 9  22 30
plate 568449 B1 7 10 23 27
plate 568449 B2 8 11 24 32
plate 568450 A1 1 12 17 29
plate 568450 A2 2 13 18 26
plate 568450 B1 3 14 19 31
plate 568450 B2 4 15 20 28
plate 568451 A1 6 16 21 32
plate 568451 A2 7 9  22 29
plate 568451 B1 8 10 23 26
plate 568451 B2 1 11 24 31
plate 568452 A1 2 12 17 28
plate 568452 A2 3 13 18 25
plate 568452 B1 4 14 19 30
plate 568452 B2 5 15 20 27
plate 568453 A1 7 16 21 31
plate 568453 A2 8 9  22 28
plate 568453 B1 1 10 23 25
plate 568453 B2 2 11 24 30
plate 568454 A1 3 12 17 27
plate 568454 A2 4 13 18 32
plate 568454 B1 5 14 19 29
plate 568454 B2 6 15 20 26
plate 568455 A1 8 16 21 30
plate 568455 A2 1 9  22 27
plate 568455 B1 2 10 23 32
plate 568455 B2 3 11 24 29
plate 568456 A1 4 12 17 26
plate 568456 A2 5 13 18 31
plate 568456 B1 6 14 19 28
plate 568456 B2 7 15 20 25
plate 568457 A1 1 16 21 29
plate 568457 A2 2 9  22 26
plate 568457 B1 3 10 23 31
plate 568457 B2 4 11 24 28
plate 568458 A1 5 12 17 25
```

plate 568458 A2 6 13 18 30
plate 568458 B1 7 14 19 27
plate 568458 B2 8 15 20 32
plate 568459 A1 2 16 21 28
plate 568459 A2 3 9 22 25
plate 568459 B1 4 10 23 30
plate 568459 B2 5 11 24 27
plate 568460 A1 6 12 17 32
plate 568460 A2 7 13 18 29
plate 568460 B1 8 14 19 26
plate 568460 B2 1 15 20 31
plate 568461 A1 4 9 22 32
plate 568461 A2 5 10 23 29
plate 568461 B1 6 11 24 26
plate 568461 B2 7 12 17 31
plate 568462 A1 8 13 18 28
plate 568462 A2 1 14 19 25
plate 568462 B1 2 15 20 30
plate 568462 B2 3 16 21 27
plate 568463 A1 5 9 22 31
plate 568463 A2 6 10 23 28
plate 568463 B1 7 11 24 25
plate 568463 B2 8 12 17 30
plate 568464 A1 1 13 18 27
plate 568464 A2 2 14 19 32
plate 568464 B1 3 15 20 29
plate 568464 B2 4 16 21 26
plate 568465 A1 6 10 22 29
plate 568465 A2 7 11 23 26
plate 568465 B1 8 12 24 31
plate 568465 B2 1 13 17 28
plate 568466 A1 2 14 18 25
plate 568466 A2 3 15 19 30
plate 568466 B1 4 16 20 27
plate 568466 B2 5 9 21 32
plate 568467 A1 7 10 22 28
plate 568467 A2 8 11 23 25
plate 568467 B1 1 12 24 30
plate 568467 B2 2 13 17 27
plate 568468 A1 3 14 18 32
plate 568468 A2 4 15 19 29
plate 568468 B1 5 16 20 26
plate 568468 B2 6 9 21 31
plate 568469 A1 8 10 22 27
plate 568469 A2 1 11 23 32

plate 568469 B1 2 12 24 29
plate 568469 B2 3 13 17 26
plate 568470 A1 4 14 18 31
plate 568470 A2 5 15 19 28
plate 568470 B1 6 16 20 25
plate 568470 B2 7 9 21 30
plate 568471 A1 1 10 22 26
plate 568471 A2 2 11 23 31
plate 568471 B1 3 12 24 28
plate 568471 B2 4 13 17 25
plate 568472 A1 5 14 18 30
plate 568472 A2 6 15 19 27
plate 568472 B1 7 16 20 32
plate 568472 B2 8 9 21 29
plate 568473 A1 3 11 23 30
plate 568473 A2 4 12 24 27
plate 568473 B1 5 13 17 32
plate 568473 B2 6 14 18 29
plate 568474 A1 7 15 19 26
plate 568474 A2 8 16 20 31
plate 568474 B1 1 9 21 28
plate 568474 B2 2 10 22 25
plate 568475 A1 4 11 23 29
plate 568475 A2 5 12 24 26
plate 568475 B1 6 13 17 31
plate 568475 B2 7 14 18 28
plate 568476 A1 8 15 19 25
plate 568476 A2 1 16 20 30
plate 568476 B1 2 9 21 27
plate 568476 B2 3 10 22 32
plate 568477 A1 5 11 23 28
plate 568477 A2 6 12 24 25
plate 568477 B1 7 13 17 30
plate 568477 B2 8 14 18 27
plate 568478 A1 1 15 19 32
plate 568478 A2 2 16 20 29
plate 568478 B1 3 9 21 26
plate 568478 B2 4 10 22 31
plate 568479 A1 6 11 23 27
plate 568479 A2 7 12 24 32
plate 568479 B1 8 13 17 29
plate 568479 B2 1 14 18 26
plate 568480 A1 2 15 19 31
plate 568480 A2 3 16 20 28
plate 568480 B1 4 9 21 25

plate 568480 B2 5 10 22 30
plate 568481 A1 7 12 23 25
plate 568481 A2 8 13 24 30
plate 568481 B1 1 14 17 27
plate 568481 B2 2 15 18 32
plate 568482 A1 3 16 19 29
plate 568482 A2 4 9 20 26
plate 568482 B1 5 10 21 31
plate 568482 B2 6 11 22 28
plate 568483 A1 8 12 23 32
plate 568483 A2 1 13 24 29
plate 568483 B1 2 14 17 26
plate 568483 B2 3 15 18 31
plate 568484 A1 4 16 19 28
plate 568484 A2 5 9 20 25
plate 568484 B1 6 10 21 30
plate 568484 B2 7 11 22 27
plate 568485 A1 2 13 24 28
plate 568485 A2 3 14 17 25
plate 568485 B1 4 15 18 30
plate 568485 B2 5 16 19 27
plate 568486 A1 6 9 20 32
plate 568486 A2 7 10 21 29
plate 568486 B1 8 11 22 26
plate 568486 B2 1 12 23 31
plate 568487 A1 3 13 24 27
plate 568487 A2 4 14 17 32
plate 568487 B1 5 15 18 29
plate 568487 B2 6 16 19 26
plate 568488 A1 7 9 20 31
plate 568488 A2 8 10 21 28
plate 568488 B1 1 11 22 25
plate 568488 B2 2 12 23 30
plate 568489 A1 4 13 24 26
plate 568489 A2 5 14 17 31
plate 568489 B1 6 15 18 28
plate 568489 B2 7 16 19 25
plate 568490 A1 8 9 20 30
plate 568490 A2 1 10 21 27
plate 568490 B1 2 11 22 32
plate 568490 B2 3 12 23 29
plate 568491 A1 5 13 24 25
plate 568491 A2 6 14 17 30
plate 568491 B1 7 15 18 27
plate 568491 B2 8 16 19 32

```
plate 568492 A1 1 9  20 29
plate 568492 A2 2 10 21 26
plate 568492 B1 3 11 22 31
plate 568492 B2 4 12 23 28
plate 568493 A1 6 13 24 32
plate 568493 A2 7 14 17 29
plate 568493 B1 8 15 18 26
plate 568493 B2 1 16 19 31
plate 568494 A1 2 9  20 28
plate 568494 A2 3 10 21 25
plate 568494 B1 4 11 22 30
plate 568494 B2 5 12 23 27
plate 568495 A1 7 13 24 31
plate 568495 A2 8 14 17 28
plate 568495 B1 1 15 18 25
plate 568495 B2 2 16 19 30
plate 568496 A1 3 9  20 27
plate 568496 A2 4 10 21 32
plate 568496 B1 5 11 22 29
plate 568496 B2 6 12 23 26
```

```
pool  463431 A1 1
pool  463431 A2 2
pool  463431 B1 3
pool  463431 B2 4
pool  463432 A1 5
pool  463432 A2 6
pool  463432 B1 7
pool  463432 B2 8
pool  463433 A1 9
pool  463433 A2 10
pool  463433 B1 11
pool  463433 B2 12
pool  463434 A1 13
pool  463434 A2 14
pool  463434 B1 15
pool  463434 B2 16
pool  463435 A1 17
pool  463435 A2 18
pool  463435 B1 19
pool  463435 B2 20
pool  463436 A1 21
pool  463436 A2 22
```

```
pool  463436 B1 23
pool  463436 B2 24
pool  463437 A1 25
pool  463437 A2 26
pool  463437 B1 27
pool  463437 B2 28
pool  463438 A1 29
pool  463438 A2 30
pool  463438 B1 31
pool  463438 B2 32
plate 813201 A1 1 9 17 32
plate 813201 A2 2 10 18 29
plate 813201 B1 3 11 19 26
plate 813201 B2 4 12 20 31
plate 813202 A1 5 13 21 28
plate 813202 A2 6 14 22 25
plate 813202 B1 7 15 23 30
plate 813202 B2 8 16 24 27
plate 813203 A1 3 9  17 30
plate 813203 A2 4 10 18 27
plate 813203 B1 5 11 19 32
plate 813203 B2 6 12 20 29
plate 813204 A1 7 13 21 26
plate 813204 A2 8 14 22 31
plate 813204 B1 1 15 23 28
plate 813204 B2 2 16 24 25
plate 813205 A1 4 9  17 29
plate 813205 A2 5 10 18 26
plate 813205 B1 6 11 19 31
plate 813205 B2 7 12 20 28
plate 813206 A1 8 13 21 25
plate 813206 A2 1 14 22 30
plate 813206 B1 2 15 23 27
plate 813206 B2 3 16 24 32
plate 813207 A1 6 10 18 25
plate 813207 A2 7 11 19 30
plate 813207 B1 8 12 20 27
plate 813207 B2 1 13 21 32
plate 813208 A1 2 14 22 29
plate 813208 A2 3 15 23 26
plate 813208 B1 4 16 24 31
plate 813208 B2 5 9  17 28
plate 813209 A1 8 10 18 31
plate 813209 A2 1 11 19 28
plate 813209 B1 2 12 20 25
```

plate 813209 B2 3 13 21 30
plate 813210 A1 4 14 22 27
plate 813210 A2 5 15 23 32
plate 813210 B1 6 16 24 29
plate 813210 B2 7 9 17 26
plate 813211 A1 1 10 18 30
plate 813211 A2 2 11 19 27
plate 813211 B1 3 12 20 32
plate 813211 B2 4 13 21 29
plate 813212 A1 5 14 22 26
plate 813212 A2 6 15 23 31
plate 813212 B1 7 16 24 28
plate 813212 B2 8 9 17 25
plate 813213 A1 3 12 19 25
plate 813213 A2 4 13 20 30
plate 813213 B1 5 14 21 27
plate 813213 B2 6 15 22 32
plate 813214 A1 7 16 23 29
plate 813214 A2 8 9 24 26
plate 813214 B1 1 10 17 31
plate 813214 B2 2 11 18 28
plate 813215 A1 5 12 19 31
plate 813215 A2 6 13 20 28
plate 813215 B1 7 14 21 25
plate 813215 B2 8 15 22 30
plate 813216 A1 1 16 23 27
plate 813216 A2 2 9 24 32
plate 813216 B1 3 10 17 29
plate 813216 B2 4 11 18 26
plate 813217 A1 6 12 19 30
plate 813217 A2 7 13 20 27
plate 813217 B1 8 14 21 32
plate 813217 B2 1 15 22 29
plate 813218 A1 2 16 23 26
plate 813218 A2 3 9 24 31
plate 813218 B1 4 10 17 28
plate 813218 B2 5 11 18 25
plate 813219 A1 8 13 20 26
plate 813219 A2 1 14 21 31
plate 813219 B1 2 15 22 28
plate 813219 B2 3 16 23 25
plate 813220 A1 4 9 24 30
plate 813220 A2 5 10 17 27
plate 813220 B1 6 11 18 32
plate 813220 B2 7 12 19 29

plate 813221 A1 2 13 20 32
plate 813221 A2 3 14 21 29
plate 813221 B1 4 15 22 26
plate 813221 B2 5 16 23 31
plate 813222 A1 6 9 24 28
plate 813222 A2 7 10 17 25
plate 813222 B1 8 11 18 30
plate 813222 B2 1 12 19 27
plate 813223 A1 3 13 20 31
plate 813223 A2 4 14 21 28
plate 813223 B1 5 15 22 25
plate 813223 B2 6 16 23 30
plate 813224 A1 7 9 24 27
plate 813224 A2 8 10 17 32
plate 813224 B1 1 11 18 29
plate 813224 B2 2 12 19 26
plate 813225 A1 5 15 21 26
plate 813225 A2 6 16 22 31
plate 813225 B1 7 9 23 28
plate 813225 B2 8 10 24 25
plate 813226 A1 1 11 17 30
plate 813226 A2 2 12 18 27
plate 813226 B1 3 13 19 32
plate 813226 B2 4 14 20 29
plate 813227 A1 7 15 21 32
plate 813227 A2 8 16 22 29
plate 813227 B1 1 9 23 26
plate 813227 B2 2 10 24 31
plate 813228 A1 3 11 17 28
plate 813228 A2 4 12 18 25
plate 813228 B1 5 13 19 30
plate 813228 B2 6 14 20 27
plate 813229 A1 8 15 21 31
plate 813229 A2 1 16 22 28
plate 813229 B1 2 9 23 25
plate 813229 B2 3 10 24 30
plate 813230 A1 4 11 17 27
plate 813230 A2 5 12 18 32
plate 813230 B1 6 13 19 29
plate 813230 B2 7 14 20 26
plate 813231 A1 2 16 22 27
plate 813231 A2 3 9 23 32
plate 813231 B1 4 10 24 29
plate 813231 B2 5 11 17 26
plate 813232 A1 6 12 18 31

plate 813232 A2 7 13 19 28
plate 813232 B1 8 14 20 25
plate 813232 B2 1 15 21 30
plate 813233 A1 4 16 22 25
plate 813233 A2 5 9 23 30
plate 813233 B1 6 10 24 27
plate 813233 B2 7 11 17 32
plate 813234 A1 8 12 18 29
plate 813234 A2 1 13 19 26
plate 813234 B1 2 14 20 31
plate 813234 B2 3 15 21 28
plate 813235 A1 5 16 22 32
plate 813235 A2 6 9 23 29
plate 813235 B1 7 10 24 26
plate 813235 B2 8 11 17 31
plate 813236 A1 1 12 18 28
plate 813236 A2 2 13 19 25
plate 813236 B1 3 14 20 30
plate 813236 B2 4 15 21 27
plate 813237 A1 7 10 23 27
plate 813237 A2 8 11 24 32
plate 813237 B1 1 12 17 29
plate 813237 B2 2 13 18 26
plate 813238 A1 3 14 19 31
plate 813238 A2 4 15 20 28
plate 813238 B1 5 16 21 25
plate 813238 B2 6 9 22 30
plate 813239 A1 1 10 23 25
plate 813239 A2 2 11 24 30
plate 813239 B1 3 12 17 27
plate 813239 B2 4 13 18 32
plate 813240 A1 5 14 19 29
plate 813240 A2 6 15 20 26
plate 813240 B1 7 16 21 31
plate 813240 B2 8 9 22 28
plate 813241 A1 2 10 23 32
plate 813241 A2 3 11 24 29
plate 813241 B1 4 12 17 26
plate 813241 B2 5 13 18 31
plate 813242 A1 6 14 19 28
plate 813242 A2 7 15 20 25
plate 813242 B1 8 16 21 30
plate 813242 B2 1 9 22 27
plate 813243 A1 4 11 24 28
plate 813243 A2 5 12 17 25

plate 813243 B1 6 13 18 30
plate 813243 B2 7 14 19 27
plate 813244 A1 8 15 20 32
plate 813244 A2 1 16 21 29
plate 813244 B1 2 9 22 26
plate 813244 B2 3 10 23 31
plate 813245 A1 6 11 24 26
plate 813245 A2 7 12 17 31
plate 813245 B1 8 13 18 28
plate 813245 B2 1 14 19 25
plate 813246 A1 2 15 20 30
plate 813246 A2 3 16 21 27
plate 813246 B1 4 9 22 32
plate 813246 B2 5 10 23 29
plate 813247 A1 7 11 24 25
plate 813247 A2 8 12 17 30
plate 813247 B1 1 13 18 27
plate 813247 B2 2 14 19 32
plate 813248 A1 3 15 20 29
plate 813248 A2 4 16 21 26
plate 813248 B1 5 9 22 31
plate 813248 B2 6 10 23 28
plate 813249 A1 1 13 17 28
plate 813249 A2 2 14 18 25
plate 813249 B1 3 15 19 30
plate 813249 B2 4 16 20 27
plate 813250 A1 5 9 21 32
plate 813250 A2 6 10 22 29
plate 813250 B1 7 11 23 26
plate 813250 B2 8 12 24 31
plate 813251 A1 3 13 17 26
plate 813251 A2 4 14 18 31
plate 813251 B1 5 15 19 28
plate 813251 B2 6 16 20 25
plate 813252 A1 7 9 21 30
plate 813252 A2 8 10 22 27
plate 813252 B1 1 11 23 32
plate 813252 B2 2 12 24 29
plate 813253 A1 4 13 17 25
plate 813253 A2 5 14 18 30
plate 813253 B1 6 15 19 27
plate 813253 B2 7 16 20 32
plate 813254 A1 8 9 21 29
plate 813254 A2 1 10 22 26
plate 813254 B1 2 11 23 31

plate 813254 B2 3 12 24 28
plate 813255 A1 6 14 18 29
plate 813255 A2 7 15 19 26
plate 813255 B1 8 16 20 31
plate 813255 B2 1 9 21 28
plate 813256 A1 2 10 22 25
plate 813256 A2 3 11 23 30
plate 813256 B1 4 12 24 27
plate 813256 B2 5 13 17 32
plate 813257 A1 8 14 18 27
plate 813257 A2 1 15 19 32
plate 813257 B1 2 16 20 29
plate 813257 B2 3 9 21 26
plate 813258 A1 4 10 22 31
plate 813258 A2 5 11 23 28
plate 813258 B1 6 12 24 25
plate 813258 B2 7 13 17 30
plate 813259 A1 1 14 18 26
plate 813259 A2 2 15 19 31
plate 813259 B1 3 16 20 28
plate 813259 B2 4 9 21 25
plate 813260 A1 5 10 22 30
plate 813260 A2 6 11 23 27
plate 813260 B1 7 12 24 32
plate 813260 B2 8 13 17 29
plate 813261 A1 3 16 19 29
plate 813261 A2 4 9 20 26
plate 813261 B1 5 10 21 31
plate 813261 B2 6 11 22 28
plate 813262 A1 7 12 23 25
plate 813262 A2 8 13 24 30
plate 813262 B1 1 14 17 27
plate 813262 B2 2 15 18 32
plate 813263 A1 5 16 19 27
plate 813263 A2 6 9 20 32
plate 813263 B1 7 10 21 29
plate 813263 B2 8 11 22 26
plate 813264 A1 1 12 23 31
plate 813264 A2 2 13 24 28
plate 813264 B1 3 14 17 25
plate 813264 B2 4 15 18 30
plate 813265 A1 6 16 19 26
plate 813265 A2 7 9 20 31
plate 813265 B1 8 10 21 28
plate 813265 B2 1 11 22 25

```
plate 813266 A1 2 12 23 30
plate 813266 A2 3 13 24 27
plate 813266 B1 4 14 17 32
plate 813266 B2 5 15 18 29
plate 813267 A1 8 9  20 30
plate 813267 A2 1 10 21 27
plate 813267 B1 2 11 22 32
plate 813267 B2 3 12 23 29
plate 813268 A1 4 13 24 26
plate 813268 A2 5 14 17 31
plate 813268 B1 6 15 18 28
plate 813268 B2 7 16 19 25
plate 813269 A1 2 9  20 28
plate 813269 A2 3 10 21 25
plate 813269 B1 4 11 22 30
plate 813269 B2 5 12 23 27
plate 813270 A1 6 13 24 32
plate 813270 A2 7 14 17 29
plate 813270 B1 8 15 18 26
plate 813270 B2 1 16 19 31
plate 813271 A1 3 9  20 27
plate 813271 A2 4 10 21 32
plate 813271 B1 5 11 22 29
plate 813271 B2 6 12 23 26
plate 813272 A1 7 13 24 31
plate 813272 A2 8 14 17 28
plate 813272 B1 1 15 18 25
plate 813272 B2 2 16 19 30
plate 813273 A1 5 11 21 30
plate 813273 A2 6 12 22 27
plate 813273 B1 7 13 23 32
plate 813273 B2 8 14 24 29
plate 813274 A1 1 15 17 26
plate 813274 A2 2 16 18 31
plate 813274 B1 3 9  19 28
plate 813274 B2 4 10 20 25
plate 813275 A1 7 11 21 28
plate 813275 A2 8 12 22 25
plate 813275 B1 1 13 23 30
plate 813275 B2 2 14 24 27
plate 813276 A1 3 15 17 32
plate 813276 A2 4 16 18 29
plate 813276 B1 5 9  19 26
plate 813276 B2 6 10 20 31
plate 813277 A1 8 11 21 27
```

plate 813277 A2 1 12 22 32
plate 813277 B1 2 13 23 29
plate 813277 B2 3 14 24 26
plate 813278 A1 4 15 17 31
plate 813278 A2 5 16 18 28
plate 813278 B1 6 9 19 25
plate 813278 B2 7 10 20 30
plate 813279 A1 2 12 22 31
plate 813279 A2 3 13 23 28
plate 813279 B1 4 14 24 25
plate 813279 B2 5 15 17 30
plate 813280 A1 6 16 18 27
plate 813280 A2 7 9 19 32
plate 813280 B1 8 10 20 29
plate 813280 B2 1 11 21 26
plate 813281 A1 4 12 22 29
plate 813281 A2 5 13 23 26
plate 813281 B1 6 14 24 31
plate 813281 B2 7 15 17 28
plate 813282 A1 8 16 18 25
plate 813282 A2 1 9 19 30
plate 813282 B1 2 10 20 27
plate 813282 B2 3 11 21 32
plate 813283 A1 5 12 22 28
plate 813283 A2 6 13 23 25
plate 813283 B1 7 14 24 30
plate 813283 B2 8 15 17 27
plate 813284 A1 1 16 18 32
plate 813284 A2 2 9 19 29
plate 813284 B1 3 10 20 26
plate 813284 B2 4 11 21 31
plate 813285 A1 7 14 23 31
plate 813285 A2 8 15 24 28
plate 813285 B1 1 16 17 25
plate 813285 B2 2 9 18 30
plate 813286 A1 3 10 19 27
plate 813286 A2 4 11 20 32
plate 813286 B1 5 12 21 29
plate 813286 B2 6 13 22 26
plate 813287 A1 1 14 23 29
plate 813287 A2 2 15 24 26
plate 813287 B1 3 16 17 31
plate 813287 B2 4 9 18 28
plate 813288 A1 5 10 19 25
plate 813288 A2 6 11 20 30

```
plate 813288 B1 7 12 21 27
plate 813288 B2 8 13 22 32
plate 813289 A1 2 14 23 28
plate 813289 A2 3 15 24 25
plate 813289 B1 4 16 17 30
plate 813289 B2 5 9  18 27
plate 813290 A1 6 10 19 32
plate 813290 A2 7 11 20 29
plate 813290 B1 8 12 21 26
plate 813290 B2 1 13 22 31
plate 813291 A1 4 15 24 32
plate 813291 A2 5 16 17 29
plate 813291 B1 6 9  18 26
plate 813291 B2 7 10 19 31
plate 813292 A1 8 11 20 28
plate 813292 A2 1 12 21 25
plate 813292 B1 2 13 22 30
plate 813292 B2 3 14 23 27
plate 813293 A1 6 15 24 30
plate 813293 A2 7 16 17 27
plate 813293 B1 8 9  18 32
plate 813293 B2 1 10 19 29
plate 813294 A1 2 11 20 26
plate 813294 A2 3 12 21 31
plate 813294 B1 4 13 22 28
plate 813294 B2 5 14 23 25
plate 813295 A1 7 15 24 29
plate 813295 A2 8 16 17 26
plate 813295 B1 1 9  18 31
plate 813295 B2 2 10 19 28
plate 813296 A1 3 11 20 25
plate 813296 A2 4 12 21 30
plate 813296 B1 5 13 22 27
plate 813296 B2 6 14 23 32
```

Following a pooling step, the DNA inserts contained within the pooled clones can amplified with the same primers used in the original GENECALLING® chemistry. Use of the original primers at this step allows direct comparison to the original GENECALLING® sizing traces as a determinant of which clones to re-array for sequencing.

Once the clones are sized, the next step in the process is selection of the clones for sequencing. The selection of clones is based upon the end purpose of the database.

The process of assigning the size, as determined by capillary electrophoresis, back

to the original clone, is known as deconvolution. Deconvolution can proceed by first preparing a representation of the sizes, *e.g.*, a trace, for each pool set of a separated sample. Bands are then identified in each trace, and sizes are assigned unique identification numbers such that bands found at the same position (with some tolerance) are grouped under the same size identification. A probability matrix is then built with clones on one axis and sizes on the other. This identifies the probability of a given clone having a given size.

In a preferred embodiment, the matrix is initialized with heuristic values, and is assembled using one, more, or all, of the following rules: a) only sizes that have bands in the traces from pools which are known to contain the given clone are considered. b) bands that are found in the fraction range associated with the clone are favored above bands outside of that range. c) Extra points are given if there is more than one such band at a given size. d) The matrix is normalized exactly to have the probability of a clone to be any size add up to one (rows add up to one). e) The matrix is then iteratively renormalized to have the probability of a size to be of any clone add up to a small number derived from band multiplicities (columns add up to expected number of clones with that size). Iteratively, rows and columns are normalized this way 5-10 times, with an exact row normalization done last. 8) The results for each clone are then written to a database.

If desired, clone selection can be use to develop custom databases, *e.g.*, custom databases based on GENECALLING® traces described in, *e.g.*, United States Patent No. 5,871,697). All of the subsequences of interest are run against a band-finder. The band-finder is a computer algorithm employed in GENECALLING® to determine the apparent size (*e.g.*, as determined by capillary electrophoresis) of all of the bands in the GENECALLING® trace for that specific subsequence.

After tables of the bands and apparent sizes for each particular subsequence are compiled, the sizing data, as determined from the multiplex sizing step in the process are compared against the results of the band-finder. Preferably, only those clones within ±0.2 base-pair of a GENECALLING® band are selected for re-array.

Clones for a sequence diversity database can be selected by parsing sized clones into 0.3 base-pair wide bins. The clones are then ordered within the bins depending upon their

probability of being correctly sized, as determined by the deconvolution software. One or more representatives from each bin are selected for re-array. The clones selected for re-array are those with the highest probability of being correctly sized.

Both methods of physically re-arraying the clones (for sequence diversity and custom GENECALLING® database) are done similarly. A liquid handling robot, such as a Tecan Genesis or Packard MultiProbe, can be used to select the appropriate wells on the clone plates and re-array them on a destination plate. For example, during the re-array, 5 uL of culture is taken from the appropriate clone well and dispensed into 50 uL of media. After the re-array, the destination plate is incubated overnight at 37°C. The inserts contained within the clones can be further analyzed, *e.g.*, by sequencing. After re-array, template preparation PCR can be performed to prepare the template for sequencing.

(iii) Partitioning based on hybridization

Screening can be performed using a variety of methods that rely on hybridization between a probe sequence or sequences and a cDNA library. Members of the library containing a homologous sequence are then removed from the library. For example, a cDNA library can be brought into contact with a prepared library of known sequence in such a way that any sequence contained within the substrate library that is complimentary to any element of the subtraction library is removed or suppressed. This method obviates re-characterizing, *e.g.*, re-sequencing, already characterized members of the cDNA population.

(iv) Amplification-associated partitioning

Partitioning can also be performed in association with amplification. In particular, partitioning can be carried out during PCR amplification of adapter-ligated cDNA fragments described above. During PCR-mediated amplification of mixtures of cDNA fragments, short fragments tend to be preferentially amplified relative to large fragments. PCR conditions can be adjusted to favor the formation of larger fragments within the PCR reaction to allow efficient preferential amplification of longer fragments.

Normally, two different primers are used in PCR amplification to prime the enzymatic activity of the polymerase at each terminus of the target sequence. Conversely, if primers with identical 5' sequences are used, there is a tendency for the fragments to form lariat or pan-handle structures, due to intra-strand hybridization, which interferes with the amplification process. Because the probability of the two ends of a polymer (*i.e.*, cDNA fragment) finding one another is inversely proportional to a fractional power of the polymer length, short fragments tend to form these lariat structures more readily than do longer ones. Accordingly, this effect is exploited in the amplification of long cDNA fragments. See U.S. Patent No. 5,565,340, whose disclosure is incorporated herein by reference, in its entirety.

Long fragment amplification can be enhanced using DNA fragments to which have been ligated long adapter sequences as described above. Amplification is dependent upon a number of factors that can alter the ratio of a linear adapter structure, which is permissive for amplification, and a lariat-loop structure, which suppresses amplifications. The equilibrium constant associated with the formation of the suppressive and the permissive structures, and, therefore, the efficiency of suppression of particular DNA fragments during PCR, is primarily a function of the following factors: (*i*) differences in melting temperature of suppressive and permissive structures; (*ii*) position of the primer sequence within the adapter; (*iii*) the length of the target DNA fragments; (*iv*) PCR primer concentration; and (*v*) primary structure.

Analysis of partitioned cDNA molecules

Partitioned cDNA molecules are next analyzed by comparing the sequences to a reference nucleic acid or nucleic acids. To facilitate analysis of partitioned cDNA molecules, they can, if not subcloned previously, be ligated into an appropriate vector and transformed into cells by any applicable method.

The reference nucleic acid or nucleic acids can be any fragment for which sufficient information is available to unambiguously identify the partitioned cDNA molecule. The reference nucleic acid or nucleic acids can therefore be part of, *e.g.*, sequence databases, or databases of other characteristics that unambiguously identify a nucleic acid. Examples of such characteristics include *e.g.*, a compilation of fragment sizes associated with specific restriction enzymes for a particular gene. In some embodiments, partitioned nucleic acids will be

sequenced. The partitioned sequences can be sequenced by any method known to the art and the resulting sequence data is analyzed by computer-based systems.

Suitable databases include publicly available databases that comprehensively record all observed DNA sequences. Such databases include, *e.g.*, GenBank from the National Center for Biotechnology Information (Bethesda, Md.), the EMBL Data Library at the European Bioinformatics Institute (Hinxton Hall, UK) and databases from the National Center for Genome Research (Santa Fe, N.Mex.). However, any database containing entries for the sequences likely to be present in such a sample to be analyzed is usable in the further steps of the computer methods. Methods of searching databases are described in detail in *e.g.*, U.S. Patent No. 5,871,697, whose disclosure is incorporated herein by reference, in its entirety.

Table 1 below summarizes the various primers and adapters disclosed herein.

Table 1

SEQ ID NO:	Name	Sequence (from 5' to 3')
1	RS	CTCTCCGATG CAGGTGGC
2	RXC	AGCACACTCC AGCCTCTCTC CGAGCACATG CGACACTGAG TACTAC
3	RXA	AGCACACTCC AGCCTCTCTC CGAGCACATG CGACACTGAG TACTAA
4	RJC	AGCACACTCC AGCCTCTCTC CGAACCGACG TCGAATATCC ATGCAGC
5	RJA	AGCACACTCC AGCCTCTCTC CGAACCGACG TCGAATATCC ATGCAGA
6	J23	ACCGACGTCG AATATCCATG CAG
7	R23	AGCACACTCC AGCCTCTCTC CGA
8	NR17	AGCACACTCC AGCCTCT
9	RA24	AGCACACTCC AGCCTCTCTC CGAA
10	RC24	AGCACACTCC AGCCTCTCTC CGAC
11	JA24	ACCGACGTCG AATATCCATG CAGA
12	JC24	ACCGACGTCG AATATCCATG CAGC
13	DT-R	AGCACACTCC AGCCTCTCTC CGA
14		AGCACACTCC AGCCTCTCTC CGATTTTTTT TTTTTTTTTT TTT

EXAMPLES

The invention will be further described in the following examples, which do not limit the scope of the invention described in the claims. Examples 1-6 collectively describe the synthesis

and amplification of cDNA subfractions enriched for the 5' terminal sequences of mRNA molecules. Example 7 describes CloneSizing™.

Example 1. 5' cDNA Synthesis—phosphatase/pyrophosphate digestion

For each reaction, 2.5 µg mRNA (do not exceed 3 µg total) is added to H_2O so as to provide a total volume of 73.5 µl. This mixture is then heated to 65°C for 10 minutes, and quick-cooled on ice. The CIAP Cocktail (see below) is made as follows:

CIAP Cocktail:

For each reaction:	10 µl 10x CIAP buffer	110 µl
	2.5 µl RNasin (Promega) x 11	27.5 µl
	10 µl 0.1 M DTT	110 µl
	4 µl 0.01 U/µl CIAP*	35 µl

1) 26.5 µl of the above enzyme mixture is added to each 3 µl mRNA to give a total volume of 30.5 µl. 73.5 µl of the RNA mix is then added to give a final volume of 100 µl.

2) Incubate at 37°C for 40 minutes.

3) Add 100 µl TE buffer (10 mM Tris pH 8.0; 0.1 mM EDTA).

4) Add 200 µl Acid-Phenol.

5) Mix vigorously.

6) Add 200 µl Chloroform-Isoamyl Alcohol (24:1 v/v).

7) Mix vigorously.

8) Centrifuge in a microfuge at maximum speed for 10 minutes.

9) Remove supernatant and transfer to new tube. Discard bottom layer.

10) Repeat steps 4-9 (only for CIAP treatment, not in later steps).

11) Add 2 µl ssDNA carrier and 20 µl 3 M Sodium Acetate to each tube.

12) Vortex 10 seconds and add 440 µl of absolute ethanol.

13) Vortex 10 seconds and incubate at least 30 minutes at –80°C.

14) Centrifuge samples at 13,200 x g for 15 minutes.

15) Wash nucleic acid pellets with 70% ethanol and air-dry pellet.

16) Dissolve nucleic acid pellet in 70 µl water and cool on ice.

17) Centrifuge for 10-15 seconds at maximum speed.

18) Transfer contents of tubes to 8-strip tubes.

19) Add 30 µl TAP cocktail (see below).

TAP Cocktail:

For each reaction:	10 µl 10x TAP buffer	110 µl
	2.5 µl RNasin x 11	27.5 µl
	15.5 µl H_2O	170.5 µl
	2.0 µl 10 U/µl TAP (Epicenter)	22 µl

20) Add 30 µl of above mixture to each 70µl CIAP-treated sample for a total volume of 100 µl.

21) Incubate at 37°C for 45 minutes.

22) Repeat Phenol/Chloroform extraction and precipitation as above in steps 6-9 and then 11-15 (do not resuspend pellet).

Example 2. 5' cDNA synthesis: DNA-RNA hybrid primer ligation

1) Transfer samples from Example 1 to 8-strip tubes.

2) Resuspend pellet in Ligation Cocktail (see below).

Ligation Cocktail:

For each reaction:		
	3 µl 10 mM ATP	33 µl
	1 µl RNasin x 11	11 µl
	4.5 µl H_2O	49.5 µl
	2 µl R-BAP-TAP DNA/RNA hybrid oligomer	22 µl

3) Add 10.5 µl of above mixture to each pellet, dissolve pellet completely at room temperature by (preferably) tapping the tube or vortexing if needed.

4) Make an enzyme mix as follows:

Enzyme Mixture:

For each reaction:		
	30 µl H_20	330 µl
	12 µl 5x DNA Ligase Buffer (Life Tech) x 11	132 µl
	1.5 µl RNasin	16.5 µl
	6 µl T_4 RNA Ligase (Life Tech.)	66 µl
	Total reaction volume 60 µl	

5) Incubate overnight at 20°C.

6) Repeat Phenol/Chloroform and precipitation as above in CIP/TAP Cocktail protocol steps 6-9 and 11-15 (do not resuspend pellet).

Example 3. 5' cDNA Synthesis: cDNA First-Strand Synthesis

1) Resuspend cDNA pellet in Random Hexamer Cocktail (see below).

Random Hexamer Cocktail:

For each reaction:		
	10 µl H_2O x 11	110 µl
	0.5 µl random hexamer (dN_6-5'-Phosphate, 100 µM)	5.5 µl
	5 µl Oligo-(dT) (dT_{30}VN-5'Phosphate, 100 µM)	55 µl

2) Add 15.5 µl of above mixture to each tube and resuspend pellet.

3) Heat at 70°C for 10 minutes and quick-cool on ice.

4) Make First-Strand Synthesis Cocktail as follows (see below).

First-Strand Synthesis Cocktail:

For each reaction:	6 µl 5x First-Strand Buffer	66 µl
	3 µl 10 mM dNTPs	33 µl
	3 µl 100 mM DTT x 11	33 µl
	1 µl RNase Inhibitor	11 µl

5) Add 13 µl of the above mixture to each 15.5 µl sample to give a total volume of 28.5 µl.

6) Incubate at 37°C for 2 minutes.

7) Add 1.5 µl SuperScript II RT to each reaction for a total volume of 30 µl.

8) Incubate at 37°C for 10 minutes.

9) Incubate at 42°C for 1 hour.

10) Incubate at 16°C.

11) Add 40 µl of the following DNA Ligase Mixture (see below) to each reaction tube for a total volume of 70 µl.

E. coli DNA Ligase Mixture:

For each reaction:	4 µl 10x *E. coli* Ligase Buffer x 11	44 µl
	33 µl H_2O	330 µl
	3 µl *E. coli* DNA Ligase (10 U/µl)	33 µl

12) Continue incubation at 16°C for 2 hours.

Example 4. 5' cDNA Synthesis: removal of non-ligated Primers

While the above 2 hour incubation described in Example 3 is progressing, prepare one Boehringer-Mannheim Quick-Spin G-50 columns per reaction as follows:

1) Mix the resin bed well by inverting the columns repeatedly.

2) Remove the top cap first, and then the bottom cap. This avoids bubble formation

and resultant poor performance of the spin-column.

3) Stand column vertically and allow to drain completely.

4) Add 0.75 ml of 10 mM Tris (pH 7.5) to the top of the bed without disturbing. If the bed becomes disturbed, pipette the solution up and down slowly to mix the bed uniformly and allow the bed to re-settle so as to form a uniform surface.

5) Stand column vertically and allow to drain completely.

6) Place the columns into a 15 ml conical centrifuge tube with the vendor's associated collector tube beneath the spin-column to collect the sample.

7) Centrifuge spin-column at 1000-1200 x g for 2 minutes.

8) Remove spin-column with a forceps and remove the tube with flow through and discard.

9) Carefully load the sample to the top center of the spin-column.

10) Wash the sample tube with 20 µl H_2O and load on the same column.

11) Place a new collection tube beneath each spin-column and centrifuge at 1000-1200 x g for 4 minutes.

12) Remove spin-columns and collect the flow-through into new, labeled tubes.

13) Total sample volume will be approximately 105 µl.

Example 5. 5' cDNA Synthesis: RNase (H, A, and T_1) Treatment

1) To each reaction described in Example 4 add Second-Strand Reaction Buffer (see below).

Second-Strand Reaction Buffer:

For each reaction:	3 µl 100 mM DTT	33 µl
	6 µl First-Strand Buffer	33 µl
	30 µl Second-Strand Buffer x 11	330 µl

6 µl H_2O	66 µl

2) Add 45 µl of the above mixture to each 105 µl sample to give a total volume of 150 µl.

3) Add 2 µl of RNase H to each sample.

4) Incubate at 37°C for 30 minutes to nick the RNA in RNA/DNA hybrids.

5) Make an RNase Mixture comprising: 22µl RNase H, 44 µl RNase Cocktail (Ambion; available as an RNase A and RNase T_1 mixture).

6) Heat samples to 95°C for 2 minutes.

7) Slow cool down to 37°C and continue incubation.

8) Add 3 µl RNase Mixture to each of the cDNAs, mix by pipetting up and down.

9) Continue incubation at 37°C for an additional 10 minutes.

10) Heat samples to 95°C for 2 minutes.

11) Slow cool down to 37°C and continue incubation.

12) Add an additional 3 µl of RNase Mixture to each of the cDNAs, mix by pipetting up and down.

13) Continue incubation at 37°C for an additional 15 minutes.

14) Repeat Phenol/Chloroform extraction and precipitation as above in steps 6-9 and then 11-15.

15) Dissolve pellet in 20 µl H_2O.

16) Remove a 5 µl aliquot for Second-Strand (see below) synthesis for producing 5'-cDNA for SEQCALLING™ Chemistry Protocol.

Example 6. Second-Strand Synthesis for Producing 5'-cDNA for SEQCALLING™ Chemistry

1) Generate PCR Mixture (see below) as follows:

PCR Mixture:

For each reaction:	5 µl 10x PCR Buffer x 11	55 µl
	1 µl 10 mM dNTPs	5.5 µl
	1 µl 10 µM R17 Primer	5.5 µl
	37.5 µl H_2O	412.5 µl
	0.5 µl Advantage Polymerase	5.5 µl

2) Add 45 µl of the above mixture to each 5 µl sample, for a total volume 50 µl.

3) Heat samples as per protocol below, making sure that the sample tubes are placed in the thermocycler only after it has reached >80°C.

94°C for 2 minutes	|	
55°C for 2 minutes	|	x 1 Cycle ONLY
72°C for 60 minutes	|	(Cycle designated KM-AD-2N)
4°C for long-term storage		

4) Warm reaction tubes to 37°C.

5) Make SAP Cocktail (see below) as follows

SAP Cocktail:

For each reaction:	12 µl 10x SAP Buffer x 11	132 µl
	5 µl H_2O	55 µl
	3 µl Shrimp Alkaline Phosphatase (SAP; 1 U/µl)	33 µl

6) Add 20 µl of SAP Cocktail to each reaction.

7) Heat to 37°C for 30 minutes.

8) Purify samples by Qiagen 96-well plate as manufacture's protocol.

9) Elute cDNAs in 100 µl 10mM Tris-HCl buffer and proceed with fluorometry.

Example 7. Dilution and amplification of restriction enzyme fragments for CloneSizing™ identification

A detailed protocol for re-amplification of restriction enzyme fragments of non-pooled samples follows:

Dilution and amplification of fragments:

Quantitative expression analysis ("QEA") solutions are obtained from GENECALLING® identification in the form of 384 well plates containing 2µl of each reaction. The fragments are reamplified in a 17 cycle PCR amplification as follows:

1) Dilute samples arrayed in a 96 well plate (Thermo-Fast 96, Marsh) with ultra pure water (Sigma) in a total volume of 100µl

2) PCR amplify 2µl of diluted samples in a 100ul reaction (1x Clonetch Buffer, Clontech; 0.4mM dNTPs, Boehringer Mannheim corp; 40pmole primers, Amitof; 0.4x Clonetech polymerase Advantage, Clontech) on a PTC-225 (MJ Research) thermocycler.

PCR program:

step 1 96°C for 5 minutes
step 2 96°C for 30 seconds
step 3 57°C for 1minute
step 4 72°C for 2 minutes
step 5 go to step 2 for 16 cycles
step 6 72°C for 10 minutes
step 7 14°C forever

To load the entire PCR amplification reaction in a MetaPhor® gel well, which contains at most 20µl, the DNA is precipitated and re-suspended to an appropriate volume as follows:.

3) Retrieve the 100µl PCR reaction and precipitate the DNA in individual Eppendorf tubes by adding DNA carrier (typically ~10% glycogen, from Amersham) and 5 volumes of 100% cold EtOH (AAPER Alcohol)

4) Vortex well, let sit on ice for 30 minutes

5) Centrifuge at 12,000 rpm for 15 minutes

6) Pour off supernatant, wash with 5 volumes of 70% cold EtOH (AAPER Alcohol)

7) Dry 15 minutes at room temperature

8) Re-suspend the pellet by adding 10µl of 10mM Tris pH 8.5 (Fisher)

9) Incubate 15 minutes at room temperature, vortex gently, store at –20°C or fractionate on MetaPhor® gel.

Example 8. Dilution and amplification of pooled restriction enzyme fragments for CloneSizing™ identification

A detailed protocol for the re-amplification of restriction enzyme fragments for pooled samples follows.

Dilution and amplification of fragments:

1) Prepare 200µl final PCR reaction with the same concentrations of the single project mix. Each tissue will be amplified with a unique set of primers.

2) Precipitate the entire PCR reaction, or 200µl, with twice the volumes as in single projects.

In order to fractionate an equal amount of DNA, the samples are first quantified. Three tissues are pooled and 2.25 µg of total DNA is fractionated.

1) Prepare a fluorometer 96 well plate (Fisher) such that DNA standard fills the first three columns and samples the rest of the plate. The DNA standard comes with the kit (PicoGreen® dsDNA Quantitation Kit 200-2000 assays). Each sample will be measured in duplicate.

2) Add 70µl of a 1/350 dilution in 1xTE (Ambion) of the samples to an equal volume of a 1/6 dilution in 1xTE (Ambion) of the PicoGreen® dye (Molecular Probes). Measure the DNA concentration in a SpectraFluor Tecan with an excitation filter of 485nm and an emission filter of 535nm.

3) Calculate the DNA concentration for each sample

4) Mix 0.75 µg of each sample

5) Bring the volume with nanopure water to 13µl total

Example 9. Electrophoresis and elution of restriction fragments in agarose gels

A Metaphor agarose gel is prepared as follows:

1) Place a 15x25 cm gel tray (BioRAD) with a 26 teeth comb 1.5x6 mm in the first slot (BioRAD)

2) In a 500ml flask place a large stir bar and add 160ml chilled (to 5-10°C) 1xTAE (BioRAD)

3) Weigh 4.8g or 6.4g to make a 3% or 4% Metaphor gel respectively

4) Slowly sprinkle, while stirring, the agarose in the solution, till all is incorporated

5) Remove the stir bar and weigh the flask on a balance (Mettler Toledo)

6) Let solution sit for 15 minutes

7) Cover the flask with a plastic wrap (Sealwrap Borden) and pierce a small hole in the plastic to allow ventilation

8) To make a 3% gel proceed to step 9. To make a 4% MetaPhor® gel steps a, b and c are required

   a. Heat flask in a microwave (Turntable microwave oven GE) on medium for 2 minutes
   b. Remove flask from microwave
   c. Let sit for 15 minutes

9) Heat flask in microwave oven on medium for 2 minutes

10) Remove flask from microwave oven

11) Gently swirl the flask to mix the agarose solution

12) Heat the flask in microwave oven on high power until solution comes to a boil

13) Hold at boiling point for 1 minute

14) Place the flask on the balance and add sufficient hot water to obtain initial weight. Mix thoroughly

15) Pour solution in the gel tray

16) Let sit at room Temperature for 30 minutes

17) Place at 4°C for 30 minutes

18) Slowly remove the comb by first flooding the teeth with nanopure water

Either non-pooled or pooled fragments are used.

Fractionation of fragments – Non-pooled:

Fragments are separated based on their size by gel electrophoresis.

1) Add 4µl of 6x Ficoll dye (0.25% Xylene cyanol, Sigma; 0.25% Bromophenol Blue, Amresco; 15% Ficoll 400, Sigma) to the eluate

2) Load half the solution, or 7µl, on each MetaPhor® gel, and on each side of the sample, a mix of two ladders (2µg of 1kb DNA ladder from Life Technologies and 400ng of Superladder from Gensura)

3) Run at constant voltage, 17V/cm, for 2 hours and 2 hours and 15 minutes for the 4% and 3% respectively in SuperSub Electrophoresis HE 100B box (Pharmacia). The gel is covered with 2 mm of re-circulating 0.5x TAE buffer (BioRAD), chilled at –4°C (Neslab RTE-100 circulator). The circulator is filled with 5 liters of 25% Ethylene Glycol (J.T.Baker)

4) After completion of the run, the gel is submerged in 200ml EtBr (Sigma) 0.5µg/m solution for 20 minutes, gently shaking

5) Then washed for 15 minutes in 200ml of 0.5x TAE (BioRAD), gently shaking

6) Both MetaPhor® gels are cut in 24 fractions with the device shown in FIGS. 5A, 5B, and 5C.

7) Fractions from 500 to 200 bp are cut from the 3% MetaPhor® gel and from 220 to 80 bp from the 4% MetaPhor® gel

8) The agarose plugs are pocked in a 96 well filter plate MANANLY10 (Millipore), such as the plate is divided in 4 sectors of 3 columns each, with every sector containing the plugs of a predetermined subsequence. Within a sector, the plugs are arranged such that the ones containing the highest molecular weight DNA are pocked in the well corresponding to the first row and column

9) Place a 96 well culture plate (Falcon 3077) underneath it with centrifuge alignment

frame (Millipore) between filter and culture plate.

Fractionation of fragments – Pooled, tagged tissues:

1) Add 3µl of 6x Ficoll dye (0.25% Xylene cyanol, Sigma; 0.25% Bromophenol Blue, Amresco; 15% Ficoll 400, Sigma) to the eluate

2) Load 16µl, or the total amount, on each MetaPhor® gel with a ladder (2µg of 1kb DNA ladder from Life Technologies and 400ng of Superladder from Gensura) between each sample

3) Run at constant voltage, 17V/cm, for 2 hours and 2 hours and 15 minutes for the 4% and 3% respectively, in SuperSub Electrophoresis HE 100B box (Pharmacia). The gel is covered with 2 mm of re-circulating 0.5x TAE buffer (BioRAD), chilled at –4°C (Neslab RTE-100 circulator). The circulator is filled with 25% Ethylene Glycol (J.T.Baker)

4) After completion of the run, the gel is submerged in 200ml of 0.5µg/ml EtBr (Sigma) solution for 20 minutes, gently shaking

5) Wash for 15 minutes in 200ml of 0.5x TAE (BioRAD), gently shaking

6) Both MetaPhor® gels are cut in 24 fractions with the device shown in FIGS. 5A, 5B, and 5C.

7) Twenty four fractions from 500 to 200 bp are cut from the 3% MetaPhor® gel and from 220 to 80 bp from the 4% MetaPhor® gel

8) The agarose plugs are pocked in a 96 well filter plate MANANLY10 (Millipore), such as the plate is divided in 4 sectors of 3 columns each, with every sector containing the plugs of a particular subsequence. Within a sector, the plugs are arranged such that the ones containing the highest molecular weight DNA are pocked in the well corresponding to the first row and column

9) Place a 96 well culture plate (Falcon 3077) underneath it with centrifuge alignment frame (Millipore) between filter and culture plate

Signature Primers:

Signature 1	5'-AgCACTCTCCAgCCTCTCACCgAC-3' (SEQ ID NO: 15) 5'-AgCACTCTCCAgCCTCTCACCgAA-3' (SEQ ID NO:16)

	5'-ACCgACgTCgACTATCCATgAAgC-3' (SEQ ID NO:17) 5'-ACCgACgTCgACTATCCATgAAgA-3' (SEQ ID NO:18)
Signature 2	5'-gggTgCATCCAgCCTCTCACCgAA-3' (SEQ ID NO:19) 5'-gggTgCATCCAgCCTCTCACCgAC-3' (SEQ ID NO:20) 5'-gggTgCATCgACTATCCATgAAga-3' (SEQ ID NO:21) 5'-gggTgCATCgACTATCCATgAAgC-3' (SEQ ID NO:22)
Signature 3	5'-ATCTCTgTCCAgCCTCTCACCgAA-3' (SEQ ID NO:23) 5'-ATCTCTgTCCAgCCTCTCACCgAC-3' (SEQ ID NO:24) 5'-ATCTCTgTCgACTATCCATgAAgA-3' (SEQ ID NO:25) 5'-ATCTCTgTCgACTATCCATgAAgC-3' (SEQ ID NO:26)
Signature 4	5'-gTTTCTCTCgACTATCCATgAAgA-3' (SEQ ID NO:27) 5'-gTTTCTCTCgACTATCCATgAAgC-3' (SEQ ID NO:28) 5'-gTTTCTCTCCAgCCTCTCACCgAA-3' (SEQ ID NO:29) 5'-gTTTCTCTCCAgCCTCTCACCgAC-3' (SEQ ID NO:30)

Elution of fragments:

The DNA is eluted by centrifuge force at 6,000 rpm for 20 minutes at room temperature, as follows:

1) Centrifuge in the Sigma-Qiagen 4-15C centrifuge at 6000 rpm (5796rcf) for 20 minutes

2) Proceed to ligation or store the plate at –20°C with aluminum tape (3M) sealing the plate

Example 10. PCR amplification of multiple pooled clones

Pooled clones are amplified using labeled primer set used in original GENECALLING® Chemistry and multiplex-sizing PCR. Primers include:

5'-FAM-ACCgACgTCgACTATCCATgAAgA-3' (SEQ ID NO:31)

5'-FAM-ACCgACgTCgACTATCCATgAAgC-3' (SEQ ID NO:32)

5'-Biot-gggTgCATCCAgCCTCTCACCgAA-3' (SEQ ID NO:33)

5'-Biot-gggTgCATCCAgCCTCTCACCgAC-3' (SEQ ID NO:34)

Pooled PCR

1. Retrieve buffer, dNTPs and primers from –20. Thaw on ice.
2. Retrieve 150mL Falcon Tube and place on ice.
3. Set up the Multidrop (Labsystems, Finland) volume to 25 and columns to 24
4. Prepare PCR cocktail according to the chart below. All reagents should be kept on ice, use filter tips and vortex the buffer, dNTPs and primers prior to adding to cocktail. Vortex or mix well again when all components have been added except polymerase. Do Not Vortex polymerase. Transport polymerase in a –20°C reagent cooler and add to cocktail last.

Pooled PCR Cocktail

Component	1 Plate	2 Plates	3 Plates	4 Plates
Sigma Water	8740.15	17480.3	26220.45	34960.6
Clonetech Buffer	1050	2100	3150	4200
5M Betaine	113.4	226.8	340.2	453.6
10mM dNTPs	457.8	915.6	1373.4	1831.2
100pmole/uL Labeled Primers	46.15	92.3	138.45	184.6
Polymerase	77.4	154.8	232.2	309.6

5. Place the Multidrop tubes into a 150mL Falcon tube, which is on ice.
6. Prime the Multidrop just until each jet is despensing PCR mix. Allowing the prime to proceed too long leave you with a shortage for the last plate
7. Transfer 25 µl of the PCR mixture to each well of the labelled 384 well polypropylene plate. The PCR mix for 4^+ plates is designed with enough wobble factor in it to account for the extra needed by the Multidrop.
8. Transfer samples from multiplexed culture plate to the labelled 384 well, using the following procedure:

Retrieve the 384 prong transferring tool ("frogger")

a) Swirl in 250mL of 10% bleach

b) Rinse twice in sterile nanopure

c) Swirl in 250mL of 85% ethanol

d) Remove ethanol and ignite the residual liquid by carefully exposing all the metal prongs to fire.

e) Allocate a moment between each step (to allow frogger to drip-dry before the next rinse) and

f) After the flaming portion of sterilization to allow adequate time for metal prongs to cool.

9. Once cool, place the tips of the frogger into the appropriate culture plate, swirl gently, lift straight up, and then transfer to the labelled 384-well polypropylene plate. Sterilize frogger for the next plate.

Thermocycler Profile: (All temperatures are in Celsius)

a. 96° for 5:00 min
b. 96° for 1:00 min.
c. 58° for 1:00 min
d. 72° for 2:00 min
e. Goto 2, 24 times
f. 72° for 10:00 min
g. 14° forever
h. End

After PCR, the product is diluted prior to sizing on a MegaBACE electrophoresis system using the following protocol.

10. Transfer 25 µl of 1X TE buffer to each well of the labelled 384 well polypropylene plate.

11. Transfer 2.5 µl of PCR product from the PCR plate to the 1X TE buffer dilution plate.

Example 11. Template Preparation PCR

1. Retrieve buffer, dNTPs and primers from –20°C. Thaw on ice.

2. Retrieve 175mL Falcon Tube Label and place on ice.

Prepare the PCR cocktail according to the chart below. Be sure to keep all reagents on ice, use filter tips and vortex the buffer, dNTPs and primers prior to adding to cocktail. Vortex or mix well again when all components have been added except polymerase.

Sequencing Template PCR Cocktail

Component	/Reaction	1 Plate(513rxn)	2 Plates(913rxn)	3 Plates	4 Plates
Sigma Water	19.32uL	9911uL	17639uL	25367uL	33095uL
Clonetech 10X Buffer	2.5uL	1283uL	2283uL	3283uL	4283uL
5M Betaine	1.0uL	513uL	913uL	1313uL	1713uL
10mMdNTPs	0.5uL	257uL	457uL	657uL	857uL
DYN-A Primer 20pmole/uL	0.5uL	257uL	457uL	657uL	857uL
DYN-A Rev Primer 20pmole/uL	0.5uL	257uL	457uL	657uL	857uL
Clonetech Polymerase	0.2uL	103uL	183uL	263uL	343uL

3. Transfer 25 µl of the PCR mixture to each well of the labeled 384 well Sequencing Template PCR plate.

4. Transfer samples from multiplexed culture plate to the labelled 384 well, using the following procedure:
   a) Retrieve the frogger (384 prong transferring tool)
   b) Swirl in 250mL of 10% bleach sterilization bath
   c) Rinse twice in sterile nanopure

d) Swirl in 250mL of 85% ethanol sterilization bath

e) Remove ethanol and ignite the residual liquid by carefully exposing all the metal prongs to fire.

f) Allocate a moment between each step (to allow frogger to drip-dry before the next rinse) and

After the flaming portion of sterilization to allow adequate time for metal prongs to cool.

5. Once cool, place the tips of the frogger into the appropriate culture plate, swirl gently, lift straight up, and then transfer to the 384-well polypropylene plate.

6. Place the sealed 384 well plate into thermocycler and execute program SQ-TP

SQ-TP Profile: (All temperatures are in Celsius)

a. 96° for 5:00 min
b. 96° for 1:00 min.
c. 57° for 1:00 min
d. 72° for 1:00 min
e. Goto 2, 29 times
f. 72° for 10:00 min
g. 14° forever
h. End

Example 12. Comparison of clone complexity with and without use of a sizing step

The effect of using a clone sizing step on the complexity, i.e., the representation of rarely transcripts, of the resulting clones, is shown in FIGS. 4A and 4B. In FIG. 4A, no sizing step was used, while CloneSizing was used in the identification of the clones shown in FIG. 4B. Shown in the figures is a comparison of the frequencies (expressed in percentage) of clones derived from transcripts present at varying levels. The outer numbers represent the prevalence of a particular clone sequenced, and the inner numbers represents the percentages of the total number of clones

sequenced that fall into this abundance class. As illustrated in FIG. 4A, the sequencing results that were obtained without the use of the sizing filter demonstrated that only a small percentage of the total number of fragments that were sequenced were included low copy number fragments (*i.e.,* singletons, duplicates, and triplicates). Specifically, singletons were found to comprise only 2% of the total number of fragments sequenced, while fragments that were present at greater than 51 copies comprised 38% of the total fragments sequenced. In contrast, as illustrated in FIG. 4B, the sequencing results that were obtained with the use of the sizing filter were enriched for clones from low abundance transcripts (*i.e.,* singletons, duplicates, and triplicates). These clones constituted approximately 33% of the total fragments sequenced. In contrast, without the use of this sizing filter, these fragments were found to only comprised a total of 8% of the sequencing results.

Equivalents

Although particular embodiments have been disclosed herein in detail, this has been done by way of example for purposes of illustration only, and is not intended to be limiting with respect to the scope of the appended claims that follow. In particular, it is contemplated by the inventor that various substitutions, alterations, and modifications may be made to the invention without departing from the spirit and scope of the invention as defined by the claims. For example, the selection of the specific tissue(s) or cell line(s) that is to be utilized in the practice of the present invention is believed to be a matter of routine for a person of ordinary skill in the art with knowledge of the embodiments described herein.